

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number  
**WO 01/87932 A2**

- (51) International Patent Classification<sup>7</sup>: **C07K 14/00**
- (21) International Application Number: PCT/US01/15048
- (22) International Filing Date: 11 May 2001 (11.05.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/203,875 12 May 2000 (12.05.2000) US  
60/207,932 30 May 2000 (30.05.2000) US
- (71) Applicant: **LEXICON GENETICS INCORPORATED**  
[US/US]; 4000 Research Forest Drive, The Woodlands, TX  
77381 (US).
- (72) Inventors: **HU, Yi**; 333 Holly Creek Court #203, The  
Woodlands, TX 77381 (US). **NEPOMNICHY, Boris**;  
905 Cypress Station Drive #1515, Houston, TX 77909  
(US). **WANG, Xioaming**; 4088 Commerical Avenue,  
Northbrook, IL 60062 (US). **WALKE, D., Wade**; 7507  
Danehill Drive, Spring, TX 77389 (US). **GERHARDT,  
Brenda**; 2123 Knollbrook Lane, Spring, TX 77373 (US).  
**TURNER, C., Alexander, Jr.**; 67 Winter Wheat Place,  
The Woodlands, TX 77381 (US).
- (74) Agents: **GARRETT, Arthur, S.** et al.; Finnegan, Hen-  
derson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street,  
N.W., Washington, DC 20005-3315 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— *without international search report and to be republished  
upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*



**WO 01/87932 A2**

(54) Title: NOVEL SEVEN TRANSMEMBRANE PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Nucleotide and amino acid sequences of several G protein coupled receptors are described.

**NOVEL SEVEN TRANSMEMBRANE PROTEINS AND POLYNUCLEOTIDES  
ENCODING THE SAME**

**DESCRIPTION OF THE INVENTION**

[001] The present application claims priority to U.S. Application Serial No. 60/203,875, filed May 12, 2000, and U.S. Application Serial No. 60/207,932, filed May 30, 2000, which are hereby incorporated by reference herein for any purpose.

**Field of the Invention**

[002] The present invention relates to the discovery, identification and characterization of novel human polynucleotides that encode membrane associated proteins and receptors. The invention encompasses the described polynucleotides, host cell expression systems, the encoded proteins, fusion proteins, polypeptides and peptides, antibodies to the encoded proteins and peptides, and genetically engineered animals that lack the disclosed genes, or over express the disclosed genes, or antagonists and agonists of the proteins, and other compounds that modulate the expression or activity of the proteins encoded by the disclosed genes that can be used for diagnosis, drug screening, clinical trial monitoring, and/or the treatment of physiological or behavioral disorders.

**Background of the Invention**

[003] Membrane receptor proteins can serve as integral components of cellular mechanisms for sensing their environment, and maintaining cellular homeostasis and function. Accordingly, membrane receptor proteins are often involved in signal transduction pathways that control cell physiology, chemical communication, and gene expression. A particularly relevant class

of membrane receptors are those typically characterized by the presence of 7 conserved transmembrane domains that are interconnected by nonconserved hydrophilic loops. Such "7TM receptors" include a superfamily of receptors known as G-protein coupled receptors (GPCRs). GPCRs are typically involved in signal transduction pathways involving G-proteins or PPG proteins. As such, the GPCR family includes many receptors that are known to serve as drug targets for therapeutic agents.

### **SUMMARY OF THE INVENTION**

[004] The present invention relates to the discovery, identification, and characterization of nucleotides that encode novel GPCRs, and the corresponding novel GPCR (NGPCR) amino acid sequences. The NGPCRs described for the first time herein are transmembrane proteins that span the cellular membrane and are involved in signal transduction after ligand binding. The described NGPCRs have structural motifs found in the 7TM receptor family. Expression of the described NGPCRs can be detected in human fetal brain, brain, pituitary, cerebellum, spinal cord, thymus, spleen, lymph node, bone marrow, trachea, lung, kidney, fetal liver, liver, prostate, testis, thyroid, adrenal gland, pancreas, salivary gland, stomach, small intestine, colon, skeletal muscle, heart, uterus, placenta, mammary gland, adipose, skin, esophagus, bladder, cervix, rectum, pericardium, ovary, fetal kidney, and fetal lung cells (SEQ ID NOS:1-5), or human heart and testis (SEQ ID NOS:6 and 7). The novel human GPCR sequences described herein encode proteins of 994, 826, and 335 amino acids in length (see respectively SEQ ID NOS: 2, 4, and 7). The described NGPCRs have multiple transmembrane regions (of

about 20-30 amino acids) characteristic of 7TM proteins as well as several predicted cytoplasmic domains.

[005] Additionally contemplated are "knockout" ES cells that have been engineered using conventional methods (see, for example, PCT Applic. No. PCT/US98/03243, filed February 20, 1998, herein incorporated by reference). A gene trapped knockout ES cell line has been produced in a murine homolog of the described human sequences. Accordingly, an additional aspect of the present invention includes knockout cells and animals having genetically engineered mutations in the gene encoding the presently described NGPCRs.

[006] The invention encompasses the nucleotide sequences presented in the Sequence Listing, host cells expressing such nucleotide sequences, and the expression products of such nucleotide sequences, and: (a) nucleotide sequences that encode mammalian homologs of the described NGPCRs, including the specifically described human NGPCRs, and the human NGPCR gene products; (b) nucleotide sequences that encode one or more portions of the NGPCRs that correspond to functional domains, and the polypeptide products specified by such nucleotide sequences, including but not limited to the novel regions of the described extracellular domain(s) ECD, one or more transmembrane domain(s) (TM) first disclosed herein, and the cytoplasmic domain(s) (CD); (c) isolated nucleotide sequences that encode mutants, engineered or naturally occurring, of the described NGPCRs in which all or a part of at least one of the domains is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including but

not limited to soluble receptors in which all or a portion of the TM is deleted, and nonfunctional receptors in which all or a portion of the CD is deleted; and (d) nucleotides that encode fusion proteins containing the coding region from an NGPCR, or one of its domains (*e.g.*, an ECD) fused to another peptide or polypeptide.

[007] The invention also encompasses agonists and antagonists of the NGPCRs, including small molecules, large molecules, mutant NGPCR proteins, or portions thereof that compete with the native NGPCR, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NGPCR (*e.g.*, antisense and ribozyme molecules, and gene or regulatory sequence replacement constructs) or to enhance the expression of the described NGPCR gene (*e.g.*, expression constructs that place the described gene under the control of a strong promoter system), and transgenic animals that express a NGPCR transgene or "knock-outs" that do not express a functional NGPCR.

[008] Further, the present invention also relates to methods for the use of the described NGPCR gene and/or NGPCR gene products for the identification of compounds that modulate, *i.e.*, act as agonists or antagonists, of NGPCR gene expression and or NGPCR gene product activity. Such compounds can be used as therapeutic agents for the treatment of various symptomatic representations of biological disorders or imbalances.

### **BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS**

[009] The Sequence Listing provides the sequences of certain described NGPCR ORFs, the amino acid sequences encoded thereby, as well as an ORF with surrounding 5' and 3' regions (SEQ ID NO:5).

### **DESCRIPTION OF CERTAIN EMBODIMENTS**

[010] The human NGPCRs, described for the first time herein, are novel receptor proteins that are expressed in human cells. The human NGPCR sequences were obtained using sequences from gene trapped human cells, genomic DNA, and cDNA clones isolated from human kidney and lymph node cDNA libraries (SEQ ID NOS:1-5), or skeletal muscle cDNA libraries were used to generate SEQ ID NOS:6-7 (Edge Biosystems, Gaithersburg, MD, and Clontech, Palo Alto, CA). The described NGPCRs are transmembrane proteins that fall within the 7TM protein family of receptors. As with other GPCRs, signal transduction is triggered when a ligand binds to the receptor. Interfering with the binding of the natural ligand, or neutralizing or removing the ligand, or interfering with its binding to a NGPCR will affect NGPCR mediated signal transduction. Because of their biological significance, 7TM, and particularly GPCR, proteins have been subjected to intense scientific/commercial scrutiny (see, for example, U.S. Application Ser. Nos. 08/820,521, filed March 19, 1997, and 08/833,226, filed April 17, 1997 both of which are herein incorporated by reference in their entirety for applications, uses, and assays involving the described NGPCRs).

[011] The invention encompasses the use of the described NGPCR nucleotides, NGPCR proteins and peptides, as well as antibodies, preferably

humanized monoclonal antibodies, or binding fragments, domains, or fusion proteins thereof, to the NGPCRs (which can, for example, act as NGPCR agonists or antagonists), antagonists that inhibit receptor activity or expression, or agonists that activate receptor activity or increase its expression in the diagnosis and treatment of disease.

[012] According to certain embodiments, the nucleotide sequences encompassed by the invention can be useful for chromosome mapping. For example, the nucleotide sequences as set forth in SEQ ID NOs 1, 3, 5, and 6 are found on chromosome 2 at 2p24.1 in the human genome. In certain embodiments, these sequences can act as highly specific probes to show which regions of the chromosome actually code for protein. Thus, in certain embodiments, these sequences can provide mapping information for protein coding regions within the 2p24.1 location of chromosome 2. In certain embodiments, these sequences allow for the identification of exons and the verification of splice junction sites. In certain embodiments, these sequences also can allow for the identification of the genomic locations for the NGPCR gene in mouse. This information is useful for creating “knock-out” mice in which the expression of this protein is abrogated.

[013] In particular, the invention encompasses NGPCR polypeptides or peptides corresponding to functional domains of NGPCR (*e.g.*, ECD, TM or CD), mutated, truncated or deleted NGPCRs (*e.g.*, NGPCRs missing one or more functional domains or portions thereof, such as,  $\Delta$ ECD,  $\Delta$ TM and/or  $\Delta$ CD), NGPCR fusion proteins (*e.g.*, a NGPCR or a functional domain of a NGPCR, such as the ECD, fused to an unrelated protein or peptide such as

an immunoglobulin constant region, *i.e.*, IgFc), nucleotide sequences encoding such products, and host cell expression systems that can produce such NGPCR products.

[014] The invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of the NGPCR, as well as compounds or nucleotide constructs that inhibit expression of a NGPCR gene (transcription factor inhibitors, antisense and ribozyme molecules, or gene or regulatory sequence replacement constructs), or promote expression of NGPCR (*e.g.*, expression constructs in which NGPCR coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, etc.). The invention also relates to host cells and animals genetically engineered to express the human NGPCRs (or mutants thereof) or to inhibit or "knock-out" expression of the animal's endogenous NGPCR genes.

[015] In certain embodiments, the NGPCR proteins or peptides, NGPCR fusion proteins, NGPCR nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NGPCRs or inappropriately expressed NGPCRs for the diagnosis of disease. In certain embodiments, the NGPCR proteins or peptides, NGPCR fusion proteins, NGPCR nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of NGPCR in the body. In



certain embodiments, the use of engineered host cells and/or animals may offer an advantage in that such systems allow not only for the identification of compounds that bind to an ECD of a NGPCR, but can also identify compounds that affect the signal transduced by an activated NGPCR.

[016] Finally, in certain embodiments, the NGPCR protein products (especially soluble derivatives such as peptides corresponding to the NGPCR ECD, or truncated polypeptides lacking one or more TM domains) and fusion protein products (especially NGPCR-Ig fusion proteins, *i.e.*, fusions of a NGPCR, or a domain of a NGPCR, *e.g.*, ECD,  $\Delta$ TM to an IgFc), antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists or agonists (including compounds that modulate signal transduction which may act on downstream targets in a NGPCR-mediated signal transduction pathway) can be used for therapy of such diseases. For example, in certain embodiments, the administration of an effective amount of soluble NGPCR ECD,  $\Delta$ TM, or an ECD-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NGPCR ECD would "mop up" or "neutralize" the endogenous NGPCR ligand, and prevent or reduce binding and receptor activation. In certain embodiments, nucleotide constructs encoding such NGPCR products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body delivering a continuous supply of a NGPCR, a NGPCR peptide, soluble ECD or  $\Delta$ TM or a NGPCR fusion protein that will "mop up" or neutralize a NGPCR ligand. In certain embodiments, nucleotide constructs encoding functional NGPCRs, mutant NGPCRs, as well as

antisense and ribozyme molecules can be used in "gene therapy" approaches for the modulation of NGPCR expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

[017] Various aspects of the invention are described in greater detail in the subsections below.

#### NGPCR POLYNUCLEOTIDES

[018] The cDNA sequences and deduced amino acid sequences of the described human NGPCRs are presented in the Sequence Listing. Two polymorphisms were identified including: a translationally silent A or G transition at the position represented by, for example, nucleotide 2091 of SEQ ID NO:1, or at nucleotide position 1,587 of SEQ ID NO:1 which can results in a ser or a gly being present at the corresponding amino acid position 529 of, for example, SEQ ID NO:2.

[019] The NGPCRs of the present invention include: (a) the human DNA sequences presented in the Sequence Listing and any additionally contemplated nucleotide sequence encoding a contiguous and functional NGPCR open reading frame (ORF) that hybridizes to a complement of the DNA sequences presented in the Sequence Listing under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3) and encodes a functionally

equivalent gene product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of DNA sequences that encode and express an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet which still encode a functionally equivalent NGPCR gene product. Functional equivalents of NGPCR include naturally occurring NGPCRs present in other species, and mutant NGPCRs whether naturally occurring or engineered. The invention also includes degenerate variants of the disclosed sequences.

[020] Additionally contemplated are polynucleotides encoding NGPCR ORFs, or their functional equivalents, encoded by polynucleotide sequences that are about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the polynucleotide sequences described in the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package using default parameters).

[021] In certain embodiments, the invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the described NGPCR nucleotide sequences. Such hybridization conditions may be highly stringent or moderately stringent (less highly stringent), as described above. In certain embodiments, in instances wherein the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules (in certain embodiments, those about 16 to about 100 base long, about 20 to about 80, or about 34 to about 45 base long, or those of at

least 300 nucleotides in length, or any variation or combination of sizes represented therein incorporating a contiguous region of sequence first disclosed in the present Sequence Listing) can be used in conjunction with the polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc. In certain embodiments, these DNA oligos will comprise at least 22 nucleotides. In certain embodiments, these DNA oligos will comprise at least 30 nucleotides.

[022] In certain embodiments, the oligonucleotides can be used singly or in chip format as hybridization probes. For example, in certain embodiments, a series of the described NGPCR oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NGPCRs. The oligonucleotides, in certain embodiments, between about 16 to about 40 (or any whole number within the stated range) nucleotides in length may partially overlap each other and/or the NGPCR sequence may be represented using oligonucleotides that do not overlap. Accordingly, in certain embodiments, the NGPCR polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 18 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences may begin at any nucleotide present within a sequence in the Sequence Listing and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation. For oligonucleotide probes, highly stringent conditions may refer, e.g., to washing in

6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos).

[023] In certain embodiments, the described oligonucleotides may encode or act as NGPCR antisense molecules, useful, for example, in NGPCR gene regulation (for and/or as antisense primers in amplification reactions of NGPCR gene nucleic acid sequences). With respect to NGPCR gene regulation, in certain embodiments, such techniques can be used to regulate biological functions. Further, in certain embodiments, such sequences may be used as part of ribozyme and/or triple helix sequences, also useful for NGPCR gene regulation.

[024] Additionally, the antisense oligonucleotides may comprise at least one modified base moiety which is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxymethylaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v),

5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

[025] The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

[026] In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

[027] In yet another embodiment, the antisense oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier *et al.*, 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue *et al.*, 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, FEBS Lett. 215:327-330).

[028] Oligonucleotides of certain embodiments of the invention may be synthesized by standard methods known in the art, e.g. by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples of certain embodiments, phosphorothioate oligonucleotides may be synthesized by the method of Stein *et al.* (1988, Nucl. Acids Res. 16:3209), methylphosphonate oligonucleotides

can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

[029] Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual (and periodic updates thereof), Cold Springs Harbor Press, N.Y.; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

[030] Alternatively, in certain embodiments, suitably labeled NGPCR nucleotide probes may be used to screen a human genomic library using appropriately stringent conditions or by PCR. The identification and characterization of human genomic clones in certain embodiments is helpful for identifying polymorphisms, determining the genomic structure of a given locus/allele, and/or designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene in certain embodiments can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (*e.g.*, splice acceptor and/or donor sites), etc., that can be used in diagnostics and pharmacogenomics.

[031] Further, a NGPCR gene homolog may be isolated from nucleic acid of the organism of interest in certain embodiments, by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of amino acid sequences within the NGPCR gene product disclosed herein. The

template for the reaction may be total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from, for example, human or non-human cell lines or tissue known or suspected to express a NGPCR gene allele.

[032] In certain embodiments, the PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NGPCR gene. In certain embodiments, the PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. In certain embodiments, the labeled fragment may be used to isolate genomic clones via the screening of a genomic library.

[033] In certain embodiments, PCR technology may also be utilized to isolate full length cDNA sequences. For example, in certain embodiments RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, one known, or suspected, to express a NGPCR gene). In certain embodiments, a reverse transcription (RT) reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. In certain embodiments, the resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, in certain embodiments, cDNA sequences



upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see *e.g.*, Sambrook *et al.*, 1989, *supra*.

[034] In certain embodiments, a cDNA of a mutant NGPCR gene can be isolated, for example, by using PCR. In certain embodiments, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying a mutant NGPCR allele, and by extending the new strand with reverse transcriptase. In certain embodiments, the second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these two primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the art. In certain embodiments, by comparing the DNA sequence of the mutant NGPCR allele to that of the normal NGPCR allele, the mutation(s) responsible for the loss or alteration of function of the mutant NGPCR gene product can be ascertained.

[035] Alternatively, in certain embodiments, a genomic library can be constructed using DNA obtained from an individual suspected of or known to carry the mutant NGPCR allele, or a cDNA library can be constructed using RNA from a tissue known, or suspected, to express the mutant NGPCR allele. A normal NGPCR gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NGPCR allele in such libraries. Clones containing the mutant NGPCR gene sequences can

then be purified and subjected to sequence analysis according to methods well known to those of skill in the art.

[036] Additionally, in certain embodiments, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known, or suspected, to express a mutant NGPCR allele in an individual suspected of or known to carry such a mutant allele. In this manner, gene products made by the putatively mutant tissue may be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal NGPCR gene product, as described, below, in Section 5.3. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor)

[037] Additionally, in certain embodiments, screening can be accomplished by screening with labeled NGPCR fusion proteins, such as, for example, AP-NGPCR or NGPCR-AP fusion proteins. In cases where a NGPCR mutation results in an expressed gene product with altered function (*e.g.*, as a result of a missense or a frameshift mutation), a polyclonal set of antibodies to NGPCR are likely to cross-react with the mutant NGPCR gene product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well known to those of skill in the art.

[038] The invention also encompasses nucleotide sequences that encode mutant NGPCRs, peptide fragments of the NGPCRs, truncated NGPCRs, and NGPCR fusion proteins. These include, but are not limited to,

nucleotide sequences encoding mutant NGPCRs described below; polypeptides or peptides corresponding to one or more ECD, TM and/or CD domains of the NGPCR or portions of these domains; truncated NGPCRs in which one or two of the domains is deleted, *e.g.*, a soluble NGPCR lacking the TM or both the TM and CD regions, or a truncated, nonfunctional NGPCR lacking all or a portion of the CD region. Nucleotides encoding fusion proteins may include, but are not limited to, full length NGPCR sequences, truncated NGPCRs, or nucleotides encoding peptide fragments of NGPCR fused to an unrelated protein or peptide, such as for example, a transmembrane sequence, which anchors the NGPCR ECD to the cell membrane; an IgFc domain which increases the stability and half life of the resulting fusion protein (*e.g.*, NGPCR-Ig) in the bloodstream; or an enzyme, fluorescent protein, luminescent protein which can be used as a marker.

[039] The invention also encompasses (a) DNA vectors that contain any of the foregoing NGPCR coding sequences and/or their complements (*i.e.*, antisense); (b) DNA expression vectors that contain any of the foregoing NGPCR coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences; and (c) genetically engineered host cells that contain any of the foregoing NGPCR coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell. As used herein, regulatory elements include but are not limited to inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include

but are not limited to the cytomegalovirus hCMV immediate early gene, regulatable, viral (particularly retroviral LTR promoters) the early or late promoters of SV40 adenovirus, the lac system, the trp system, the tet system, the TAC system, the TRC system, the major operator and promoter regions of phage A, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast  $\alpha$ -mating factors.

### NGPCR PROTEINS AND POLYPEPTIDES

[040] NGPCR proteins, polypeptides and peptide fragments, mutated, truncated or deleted forms of the NGPCR and/or NGPCR fusion proteins can be prepared for a variety of uses, including but not limited to the generation of antibodies, as reagents in diagnostic assays, the identification of other cellular gene products related to a NGPCR, as reagents in assays for screening for compounds that can be used as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders (*i.e.*, heartbeat rate, improper blood pressure, etc.) and disease.

[041] The Sequence Listing discloses the amino acid sequences encoded by certain NGPCR genes. The NGPCRs have initiator methionines in DNA sequence contexts consistent with translation initiation sites, followed by hydrophobic signal sequences typical of membrane associated proteins. The sequence data presented herein indicate that alternatively spliced forms of the NGPCRs exist (which may or may not be tissue specific).

[042] The NGPCR sequences of the invention include the nucleotide and amino acid sequences presented in the Sequence Listing as well as

analogues and derivatives thereof. Further, corresponding NGPCR homologues from other species are encompassed by the invention. In fact, any NGPCR protein encoded by the NGPCR nucleotide sequences described above are within the scope of the invention, as are any novel polynucleotide sequences encoding all or any novel portion of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well known, and, accordingly, each amino acid presented in the Sequence Listing, is generically representative of the well known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell *et al.* eds., Scientific American Books, New York, NY, herein incorporated by reference) are generically representative of all the various permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

[043] The invention also encompasses proteins that are functionally equivalent to the NGPCR encoded by the described nucleotide sequences as judged by any of a number of criteria, including but not limited to the ability to bind a ligand for a NGPCR, the ability to affect an identical or complementary signal transduction pathway, a change in cellular metabolism (e.g., ion flux, tyrosine phosphorylation, etc.) or a change in phenotype when the NGPCR equivalent is present in an appropriate cell type (such as the amelioration, prevention or delay of a biochemical, biophysical, or overt phenotype. Such

functionally equivalent NGPCR proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NGPCR nucleotide sequences described above but which result in a silent change, thus producing a functionally equivalent gene product. For example, in certain embodiments, one can employ a conservative amino acid substitution or substitutions which may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

[044] While random mutations can be made to NGPCR DNA (using random mutagenesis techniques well known to those skilled in the art) and the resulting mutant NGPCRs tested for activity, in certain embodiments, site-directed mutations of the NGPCR coding sequence can be engineered (using site-directed mutagenesis techniques well known to those skilled in the art) to generate mutant NGPCRs with increased function, *e.g.*, higher binding affinity for the target ligand, and/or greater signaling capacity; or decreased function, and/or decreased signal transduction capacity. In certain embodiments, one starting point for such analysis is by aligning the disclosed human sequences with corresponding gene/protein sequences from, for example, other

mammals in order to identify amino acid sequence motifs that are conserved between different species. In certain embodiments, non-conservative changes can be engineered at variable positions to alter function, signal transduction capability, or both. Additionally, where alteration of function is desired, in certain embodiments, deletion or non-conservative alterations of the conserved regions (*i.e.*, identical amino acids) can be engineered.

[045] In certain embodiments, polypeptides of the invention will be at least 85% identical, or at least 95% identical, to the polypeptides described in SEQ ID NOs 2, 4, and 7. Percent sequence identity can be determined by standard methods that are commonly used to compare the similarity in position of the amino acids of two polypeptides, to generate an optimal alignment of two respective sequences. By way of illustration, using a computer program such as BLAST or FASTA, two polypeptides are aligned for optimal matching of their respective amino acids (either along the full length of one or both sequences, or along a pre-determined portion of one or both sequences). The programs provide a "default" opening penalty and a "default" gap penalty, and a scoring matrix such as PAM 250. A standard scoring matrix can be used in conjunction with the computer program; see Dayhoff et al., in Atlas of Protein Sequence and Structure, Volume 5, Supplement 3 (1978), which is incorporated by reference herein for any purpose. In certain embodiments, the substitutions will be conservative so as to have little or no effect on the overall net charge, polarity, or hydrophobicity of the protein.

[046] In certain embodiments, the described NGPCR polynucleotide sequences can be used in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences using, for example, polynucleotide shuffling or related methodologies. Such approaches are described in U.S. Patents Nos. 5,830,721 and 5,837,458 which are herein incorporated by reference herein in their entirety for any purpose.

[047] In certain embodiments, the described sequences can be used for engineering of constitutively "on" variants for use in cell assays and genetically engineered animals using the methods and applications described in U.S. Patent Applications Ser Nos. 60/110,906, 60/106,300, 60/094,879, and 60/121,851 all of which are incorporated by reference herein in their entirety for any purpose.

[048] In certain embodiments, mutations to the NGPCR coding sequence can be made to generate NGPCRs that are better suited for expression, scale up, etc. in the host cells chosen. For example, in certain embodiments, cysteine residues can be deleted or substituted with another amino acid in order to eliminate disulfide bridges; in certain embodiments, N-linked glycosylation sites can be altered or eliminated to achieve, for example, expression of a homogeneous product that is more easily recovered and purified from yeast hosts which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid substitutions at one or both of the first or third amino acid positions of any one or more of the glycosylation recognition sequences which occur in the ECD (N-X-S or N-X-T), and/or an amino acid



deletion at the second position of any one or more such recognition sequences in the ECD will prevent glycosylation of the NGPCR at the modified tripeptide sequence. (See, *e.g.*, Miyajima *et al.*, 1986, EMBO J. 5(6):1193-1197).

[049] Peptides corresponding to one or more domains of the NGPCR (*e.g.*, ECD, TM, CD, etc.), truncated or deleted NGPCRs (*e.g.*, NGPCR in which a ECD, TM and/or CD is deleted) as well as fusion proteins in which a full length NGPCR, a NGPCR peptide, or truncated NGPCR is fused to an unrelated protein, are also within the scope of the invention and can be designed on the basis of the presently disclosed NGPCR nucleotide and NGPCR amino acid sequences. Such fusion proteins include, but are not limited to, IgFc fusions which stabilize the NGPCR protein or peptide and prolong half-life *in vivo*; or fusions to any amino acid sequence that allows the fusion protein to be anchored to the cell membrane, allowing an ECD to be exhibited on the cell surface; or fusions to an enzyme, fluorescent protein, or luminescent protein which provide a marker function.

[050] While the NGPCR polypeptides and peptides can be chemically synthesized (*e.g.*, see Creighton, 1983, *Proteins: Structures and Molecular Principles*, W.H. Freeman & Co., N.Y.), in certain embodiments, large polypeptides derived from a NGPCR and full length NGPCRs can be advantageously produced by recombinant DNA technology using techniques well known in the art for expressing nucleic acid sequences containing NGPCR gene sequences and/or coding sequences. In certain embodiments, such methods can be used to construct expression vectors containing a

presently described NGPCR nucleotide sequence and appropriate transcriptional and translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. See, for example, the techniques described in Sambrook *et al.*, 1989, *supra*, and Ausubel *et al.*, 1989, *supra*. In certain embodiments, RNA corresponding to all or a portion of a transcript encoded by a NGPCR nucleotide sequence may be chemically synthesized using, for example, synthesizers. See, for example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, M.J. ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety.

[051] According to various embodiments, a variety of host-expression vector systems may be utilized to express the NGPCR nucleotide sequences of the invention. In certain embodiments, where the NGPCR peptide or polypeptide is a soluble derivative (*e.g.*, NGPCR peptides corresponding to an ECD; truncated or deleted NGPCR in which a TM and/or CD are deleted), the peptide or polypeptide can be recovered from the culture, *i.e.*, from the host cell in cases where the NGPCR peptide or polypeptide is not secreted, and from the culture media in cases where the NGPCR peptide or polypeptide is secreted by the cells. However, such expression systems also encompass engineered host cells that express a NGPCR, or functional equivalent, *in situ*, *i.e.*, anchored in the cell membrane. In certain embodiments, purification or enrichment of NGPCR from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well known to those skilled in the art. However, in certain embodiments, such engineered

host cells themselves may be used in situations where it is important not only to retain the structural and functional characteristics of the NGPCR, but to assess biological activity, *e.g.*, in drug screening assays.

[052] In certain embodiments, the expression systems that may be used for purposes of the invention include but are not limited to microorganisms such as bacteria (*e.g.*, *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NGPCR nucleotide sequences; yeast (*e.g.*, *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NGPCR nucleotide sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing NGPCR sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing NGPCR nucleotide sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

[053] In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NGPCR gene product being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions of NGPCR protein or for raising antibodies to a NGPCR protein, for example,

vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, 1983, EMBO J. 2:1791), in which a NGPCR coding sequence may be ligated individually into the vector in frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). Typically, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

[054] In an insect system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) can be used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. A NGPCR gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of NGPCR gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the

inserted gene is expressed (e.g., see Smith *et al.*, 1983, J. Virol. 46: 584; Smith, U.S. Patent No. 4,215,051).

[055] In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the NGPCR nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NGPCR gene product in infected hosts (e.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be used for efficient translation of inserted NGPCR nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NGPCR gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, one may not employ additional translational control signals. However, in cases where only a portion of a NGPCR coding sequence is inserted, in certain embodiments, exogenous translational control signals, including, e.g., the ATG initiation codon, may be provided. Furthermore, the initiation codon typically is in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of

appropriate transcription enhancer elements, transcription terminators, etc. (See Bitter et al., 1987, Methods in Enzymol. 153:516-544).

[056] In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, and WI38 cell lines.

[057] For long-term, high-yield production of recombinant proteins, one can use stable expression. For example, cell lines which stably express the NGPCR sequences described above may be engineered. In certain embodiments, rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a

selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the NGPCR gene product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the NGPCR gene product.

[058] A number of selection systems can be used, including, but not limited to, the herpes simplex virus thymidine kinase (Wigler, *et al.*, 1977, *Cell* 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, *Proc. Natl. Acad. Sci. USA* 48:2026), and adenine phosphoribosyltransferase (Lowy, *et al.*, 1980, *Cell* 22:817) genes can be employed in tk<sup>-</sup>, hgprt<sup>-</sup> or apt<sup>-</sup> cells, respectively. Also, in certain embodiments, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler, *et al.*, 1980, *Natl. Acad. Sci. USA* 77:3567; O'Hare, *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, *et al.*, 1981, *J. Mol. Biol.* 150:1); and hygromycin (Santerre, *et al.*, 1984, *Gene* 30:147).

[059] In certain embodiments, a fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht *et al.* allows for the ready

purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88: 8972-8976). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag having six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni<sup>2+</sup>-nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

[060] In certain embodiments, NGPCR gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea pigs, rodents, pigs, micro-pigs, birds, goats, farm animals, and non-human primates, e.g., baboons, monkeys, and chimpanzees may be used in various embodiments to generate NGPCR transgenic animals.

[061] Any technique known in the art may be used to introduce a NGPCR transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe, P.C. and Wagner, T.E., 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci., USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell 57:717-723); etc. For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl.



Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

[062] The present invention provides for transgenic animals that carry the NGPCR transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals or somatic cell transgenic animals. The transgene may be integrated as a single transgene or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89:6232-6236. The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

[063] When it is desired that a NGPCR transgene be integrated into the chromosomal site of the endogenous NGPCR gene, in certain embodiments, gene targeting may be used. Briefly, in certain embodiments, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous NGPCR gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous NGPCR gene (*i.e.*, "knockout" animals).

[064] In certain embodiments, the transgene can also be selectively introduced into a particular cell type, thus inactivating the endogenous NGPCR gene in only that cell type, by following, for example, the teaching of Gu *et al.*, 1994, Science, 265:103-106. The regulatory sequences for such a

cell-type specific inactivation typically will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

[065] Once transgenic animals have been generated, the expression of the recombinant NGPCR gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of NGPCR gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the NGPCR transgene product.

#### ANTIBODIES TO NGPCR PROTEINS

[066] Antibodies that specifically recognize one or more epitopes of a NGPCR, or epitopes of conserved variants of a NGPCR, or peptide fragments of a NGPCR are also encompassed by the invention. Such antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')<sub>2</sub> fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

[067] In certain embodiments, the antibodies of the invention may be used, for example, in the detection of NGPCR in a biological sample and may,

therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of NGPCR. In certain embodiments, antibodies may be utilized in conjunction with, for example, compound screening schemes, as described below, for the evaluation of the effect of test compounds on expression and/or activity of a NGPCR gene product. In certain embodiments antibodies can be used in conjunction gene therapy to, for example, evaluate the normal and/or engineered NGPCR-expressing cells prior to their introduction into the patient. In certain embodiments, antibodies may additionally be used as a method for the inhibition of abnormal NGPCR activity. In certain embodiments, antibodies may, therefore, be utilized as part of weight disorder treatment methods.

[068] For the production of antibodies, various host animals may be immunized by injection with the NGPCR, an NGPCR peptide (e.g., one corresponding to a functional domain of the receptor, such as an ECD, TM or CD), truncated NGPCR polypeptides (NGPCR in which one or more domains, e.g., a TM or CD, has been deleted), functional equivalents of the NGPCR or mutants of the NGPCR. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic-polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal

antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

[069] Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495-497; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor *et al.*, 1983, *Immunology Today* 4:72; Cole *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80:2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

[070] In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, 1984, *Proc. Natl. Acad. Sci.*, 81:6851-6855; Neuberger *et al.*, 1984, *Nature*, 312:604-608; Takeda *et al.*, 1985, *Nature*, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Such technologies are

described in U.S. Patents Nos. 6,075,181 and 5,877,397 and their respective disclosures which are incorporated by reference herein in their entirety.

[071] In certain embodiments, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, *Science* 242:423-426; Huston *et al.*, 1988, *Proc. Natl. Acad. Sci. USA* 85:5879-5883; and Ward *et al.*, 1989, *Nature* 341:544-546) can be adapted to produce single chain antibodies against NGPCR gene products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

[072] Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, *Science*, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

[073] Antibodies to a NGPCR can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" a given NGPCR, using techniques well known to those skilled in the art. (See, e.g., Greenspan & Bona, 1993, *FASEB J* 7(5):437-444; and Nissinoff, 1991, *J. Immunol.* 147(8):2429-2438). For example antibodies which bind to a NGPCR ECD and competitively inhibit the binding of a ligand of NGPCR can be used to generate anti-idiotypes that "mimic" a NGPCR ECD and, therefore, bind and neutralize a ligand. Such

neutralizing anti-idiotypes or Fab fragments of such anti-idiotypes can be used in therapeutic regimens involving the NGPCR signaling pathway.

#### DIAGNOSIS OF ABNORMALITIES RELATED TO A NGPCR

[074] A variety of methods can be employed for the diagnostic and prognostic evaluation of disorders related to NGPCR function, and for the identification of subjects having a predisposition to such disorders.

[075] Such methods can, for example, utilize reagents such as the NGPCR nucleotide sequences and NGPCR antibodies described herein. Specifically, such reagents may be used, for example, for: (1) the detection of the presence of NGPCR gene mutations, or the detection of either over- or under-expression of NGPCR mRNA relative to a given phenotype; (2) the detection of either an over- or an under-abundance of NGPCR gene product relative to a given phenotype; and (3) the detection of perturbations or abnormalities in the signal transduction pathway mediated by NGPCR.

[076] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific NGPCR nucleotide sequence or NGPCR antibody reagent described herein, which may be conveniently used, e.g., in clinical settings, to diagnose patients exhibiting body weight disorder abnormalities.

[077] For the detection of NGPCR mutations, any nucleated cell can be used as a starting source for genomic nucleic acid. For the detection of NGPCR gene expression or NGPCR gene products, any cell type or tissue in which the NGPCR gene is expressed, such as, for example, kidney, stomach or brain cells can be utilized.

[078] Nucleic acid-based detection techniques and peptide detection techniques are described below.

#### DETECTION OF NGPCR GENES AND TRANSCRIPTS

[079] Mutations within a NGPCR gene can be detected by utilizing a number of techniques. Nucleic acid from any nucleated cell can be used as the starting point for such assay techniques, and may be isolated according to standard nucleic acid preparation procedures which are well known to those of skill in the art.

[080] DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving NGPCR gene structure, including point mutations, insertions, deletions and chromosomal rearrangements. Such assays may include, but are not limited to, Southern analyses, single stranded conformational polymorphism analyses (SSCP), and PCR analyses.

[081] Such diagnostic methods for the detection of NGPCR gene-specific mutations can involve for example, contacting and incubating nucleic acids including recombinant DNA molecules, cloned genes or degenerate variants thereof, obtained from a sample, *e.g.*, derived from a patient sample or other appropriate cellular source, with one or more labeled nucleic acid reagents including recombinant DNA molecules, cloned genes or degenerate variants thereof, as described herein, under conditions favorable for the specific annealing of these reagents to their complementary sequences within a given NGPCR gene. In certain embodiments, the lengths of these nucleic acid reagents are at least 15 to 30 nucleotides. In certain embodiments, after

incubation, all non-annealed nucleic acids are removed from the nucleic acid:NGPCR molecule hybrid. The presence of nucleic acids which have hybridized, if any such molecules exist, is then detected. Using such a detection scheme, the nucleic acid from the cell type or tissue of interest can be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtiter plate or polystyrene beads. In this case, after incubation, non-annealed, labeled nucleic acid reagents of the type described herein are easily removed. Detection of the remaining, annealed, labeled NGPCR nucleic acid reagents is accomplished using standard techniques well-known to those in the art. The NGPCR gene sequences to which the nucleic acid reagents have annealed can be compared to the annealing pattern expected from a normal NGPCR gene sequence in order to determine whether a NGPCR gene mutation is present.

[082] Alternative diagnostic methods for the detection of NGPCR gene specific nucleic acid molecules, in patient samples or other appropriate cell sources, may involve their amplification, *e.g.*, by PCR (the experimental embodiment set forth in Mullis, K.B., 1987, U.S. Patent No. 4,683,202), followed by the detection of the amplified molecules using techniques well known to those of skill in the art. The resulting amplified sequences can be compared to those which would be expected if the nucleic acid being amplified contained only normal copies of a NGPCR gene in order to determine whether a NGPCR gene mutation exists.

[083] Additionally, well-known genotyping techniques can be performed to identify individuals carrying NGPCR gene mutations. Such



techniques include, for example, the use of restriction fragment length polymorphisms (RFLPs), which involve sequence variations in one of the recognition sites for the specific restriction enzyme used.

[084] Additionally, improved methods for analyzing DNA polymorphisms which can be utilized for the identification of NGPCR gene mutations have been described which capitalize on the presence of variable numbers of short, tandemly repeated DNA sequences between the restriction enzyme sites. For example, Weber (U.S. Pat. No. 5,075,217, which is incorporated herein by reference in its entirety) describes a DNA marker based on length polymorphisms in blocks of (dC-dA)<sub>n</sub>-(dG-dT)<sub>n</sub> short tandem repeats. The average separation of (dC-dA)<sub>n</sub>-(dG-dT)<sub>n</sub> blocks is estimated to be 30,000-60,000 bp. Markers which are so closely spaced exhibit a high frequency co-inheritance, and are extremely useful in the identification of genetic mutations, such as, for example, mutations within a given NGPCR gene, and the diagnosis of diseases and disorders related to NGPCR mutations.

[085] Also, Caskey *et al.* (U.S. Pat. No. 5,364,759, which is incorporated herein by reference in its entirety) describe a DNA profiling assay for detecting short tri and tetra nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the NGPCR gene, amplifying the extracted DNA, and labeling the repeat sequences to form a genotypic map of the individual's DNA.

[086] The level of NGPCR gene expression can also be assayed by detecting and measuring NGPCR transcription. For example, RNA from a cell

type or tissue known, or suspected to express the NGPCR gene may be isolated and tested utilizing hybridization or PCR techniques such as are described, above. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a step in the assessment of cells to be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of the NGPCR gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of the NGPCR gene, including activation or inactivation of NGPCR gene expression.

[087] In certain embodiments of such a detection scheme, cDNAs are synthesized from the RNAs of interest (*e.g.*, by reverse transcription of the RNA molecule into cDNA). A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method can be chosen from among the NGPCR nucleic acid reagents described herein. In certain embodiments, the lengths of such nucleic acid reagents are at least 9-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining, by utilizing any other suitable nucleic acid staining method, or by sequencing.

[088] Additionally, it is possible to perform such NGPCR gene expression assays "*in situ*", *i.e.*, directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents such as those described above may be used as probes and/or primers for such *in situ* procedures (See, for example, Nuovo, G.J., 1992, "PCR In Situ Hybridization: Protocols And Applications", Raven Press, NY).

[089] Alternatively, if a sufficient quantity of the appropriate cells can be obtained, standard Northern analysis can be performed to determine the level of NGPCR mRNA expression.

#### DETECTION OF NGPCR GENE PRODUCTS

[090] Antibodies directed against wild type or mutant NGPCR gene products or conserved variants or peptide fragments thereof, which are discussed above, may also be used as diagnostics and prognostics, as described herein. Such diagnostic methods, may be used to detect abnormalities in the level of NGPCR gene expression, or abnormalities in the structure and/or temporal, tissue, cellular, or subcellular location of the NGPCR, and may be performed *in vivo* or *in vitro*, such as, for example, on biopsy tissue.

[091] For example, in certain embodiments, antibodies directed to epitopes of the NGPCR ECD can be used *in vivo* to detect the pattern and level of expression of the NGPCR in the body. Such antibodies can be labeled, *e.g.*, with a radio-opaque or other appropriate compound and injected into a subject in order to visualize binding to the NGPCR expressed in the

body using methods such as X-rays, CAT-scans, or MRI. Labeled antibody fragments, *e.g.*, the Fab or single chain antibody comprising the smallest portion of the antigen binding region, are used for this purpose in certain embodiments to promote crossing the blood-brain barrier and permit labeling NGPCRs expressed in the brain.

[092] Additionally, any NGPCR fusion protein or NGPCR conjugated protein whose presence can be detected, can be administered. For example, NGPCR fusion or conjugated proteins labeled with a radio-opaque or other appropriate compound can be administered and visualized *in vivo*, as discussed, above for labeled antibodies. Further such NGPCR fusion proteins as AP-NGPCR or NGPCR-Ap fusion proteins can be utilized for *in vitro* diagnostic procedures.

[093] In certain embodiments, immunoassays or fusion protein detection assays, as described above, can be utilized on biopsy and autopsy samples *in vitro* to permit assessment of the expression pattern of the NGPCR. Such assays are not confined to the use of antibodies that define a NGPCR ECD, but can include the use of antibodies directed to epitopes of any of the domains of a NGPCR, *e.g.*, the ECD, the TM and/or CD. In certain embodiments, the use of each or all of these labeled antibodies will yield useful information regarding translation and intracellular transport of the NGPCR to the cell surface, and can identify defects in processing.

[094] The tissue or cell type to be analyzed will typically include those which are known, or suspected, to express the NGPCR gene. The protein isolation methods employed herein may, for example, be such as those

described in Harlow and Lane (Harlow, E. and Lane, D., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), which is incorporated herein by reference in its entirety for any purpose. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a step in the assessment of cells that could be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of a NGPCR gene.

[095] For example, antibodies, or fragments of antibodies, such as those described herein, useful in the present invention may be used to quantitatively or qualitatively detect the presence of NGPCR gene products or conserved variants or peptide fragments thereof. This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below, this Section) coupled with light microscopic, flow cytometric, or fluorimetric detection. In certain embodiments, such techniques are used if such NGPCR gene products are expressed on the cell surface.

[096] The antibodies (or fragments thereof) or NGPCR fusion or conjugated proteins useful in the present invention may, in certain embodiments, be employed histologically, as in immunofluorescence, immunoelectron microscopy or non-immuno assays, for *in situ* detection of NGPCR gene products or conserved variants or peptide fragments thereof, or for NGPCR binding (in the case of labeled NGPCR ligand fusion protein).

[097] *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody

or fusion protein of the present invention. In certain embodiments, the antibody (or fragment) or fusion protein is applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure in certain embodiments, it is possible to determine not only the presence of a NGPCR gene product, or conserved variants or peptide fragments, or NGPCR binding, but also its distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

[098] In certain embodiments, immunoassays and non-immunoassays for NGPCR gene products or conserved variants or peptide fragments thereof will typically comprise incubating a sample, such as a biological fluid, a tissue extract, freshly harvested cells, or lysates of cells which have been incubated in cell culture, in the presence of a detectably labeled antibody capable of identifying NGPCR gene products or conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art.

[099] In certain embodiments, the biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support which is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled NGPCR antibody or NGPCR ligand fusion protein. The solid phase support may then be washed with the buffer a second time to remove unbound antibody or

fusion protein. The amount of bound label on solid support may then be detected by conventional means.

[0100] By "solid phase support or carrier" is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include, but are not limited to, glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material can have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include, but are not limited to, polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

[0101] The binding activity of a given lot of NGPCR antibody or NGPCR ligand fusion protein may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

[0102] With respect to antibodies, one of the ways in which the NGPCR antibody can be detectably labeled is by linking the same to an enzyme and

use in an enzyme immunoassay (EIA) (Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)", 1978, Diagnostic Horizons 2:1-7, Microbiological Associates Quarterly Publication, Walkersville, MD); Voller, A. *et al.*, 1978, J. Clin. Pathol. 31:507-520; Butler, J.E., 1981, Meth. Enzymol. 73:482-523; Maggio, E. (ed.), 1980, Enzyme Immunoassay, CRC Press, Boca Raton, FL.; Ishikawa, E. *et al.*, (eds.), 1981, Enzyme Immunoassay, Kigaku Shoin, Tokyo). The enzyme that is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

[0103] Detection may also be accomplished using any of a variety of other immunoassays. For example, in certain embodiments, by radioactively labeling the antibodies or antibody fragments, it is possible to detect NGPCR through the use of a radioimmunoassay (RIA) (see, for example, Weintraub,



B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

[0104] It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

[0105] In certain embodiments, the antibody can also be detectably labeled using fluorescence emitting metals such as  $^{152}\text{Eu}$ , or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

[0106] In certain embodiments, the antibody can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, thermotropic acridinium ester, imidazole, acridinium salt and oxalate ester.

[0107] In certain embodiments, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Bioluminescent compounds for purposes of labeling according to certain embodiments are luciferin, luciferase and aequorin.

#### SCREENING ASSAYS FOR COMPOUNDS THAT MODULATE NGPCR EXPRESSION OR ACTIVITY

[0108] The following assays are designed to identify compounds that interact with (*e.g.*, bind to) NGPCRs (including, but not limited to an ECD or CD of a NGPCR), compounds that interact with (*e.g.*, bind to) intracellular proteins that interact with NGPCR (including but not limited to the TM and CD of NGPCR), compounds that interfere with the interaction of NGPCR with transmembrane or intracellular proteins involved in NGPCR-mediated signal transduction, and to compounds which modulate the activity of NGPCR gene (*i.e.*, modulate the level of NGPCR gene expression) or modulate the level of NGPCR. In certain embodiments, assays may be utilized which identify compounds which bind to NGPCR gene regulatory sequences (*e.g.*, promoter sequences) and which may modulate NGPCR gene expression. See *e.g.*, Platt, K.A., 1994, J. Biol. Chem. 269:28558-28562, which is incorporated herein by reference in its entirety.

[0109] The compounds that can be screened in accordance with certain embodiments of the invention include, but are not limited to, peptides,

antibodies and fragments thereof, and other organic compounds (*e.g.*, peptidomimetics) that bind to an ECD of a NGPCR and either mimic the activity triggered by the natural ligand (*i.e.*, agonists) or inhibit the activity triggered by the natural ligand (*i.e.*, antagonists); as well as peptides, antibodies or fragments thereof, and other organic compounds that mimic the ECD of the NGPCR (or a portion thereof) and bind to and "neutralize" the natural ligand.

[0110] Such compounds may include, but are not limited to, peptides such as, for example, soluble peptides, including but not limited to members of random peptide libraries; (see, *e.g.*, Lam, K.S. *et al.*, 1991, *Nature* 354:82-84; Houghten, R. *et al.*, 1991, *Nature* 354:84-86), and combinatorial chemistry-derived molecular library made of D- and/or L- configuration amino acids, phosphopeptides (including, but not limited to members of random or partially degenerate, directed phosphopeptide libraries; see, *e.g.*, Songyang, Z. *et al.*, 1993, *Cell* 72:767-778), antibodies (including, but not limited to, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')<sub>2</sub> and FAb expression library fragments, and epitope-binding fragments thereof), and small organic or inorganic molecules.

[0111] Other compounds which can be screened in accordance with certain embodiments of the invention include but are not limited to small organic molecules that are able to cross the blood-brain barrier, gain entry into an appropriate cell (*e.g.*, in the cerebellum, the hypothalamus, etc.) and affect the expression of a NGPCR gene or some other gene involved in the NGPCR signal transduction pathway (*e.g.*, by interacting with the regulatory

region or transcription factors involved in gene expression); or such compounds that affect the activity of the NGPCR (e.g., by inhibiting or enhancing the enzymatic activity of a CD) or the activity of some other intracellular factor involved in the NGPCR signal transduction pathway.

[0112] Computer modeling and searching technologies permit identification of compounds, or the improvement of already identified compounds, that can modulate NGPCR expression or activity. Having identified such a compound or composition, the active sites or regions are identified. Such active sites might typically be ligand binding sites. The active site can be identified using methods known in the art including, for example, from the amino acid sequences of peptides, from the nucleotide sequences of nucleic acids, or from study of complexes of the relevant compound or composition with its natural ligand. In the latter case, chemical or X-ray crystallographic methods can be used to find the active site by finding where on the factor the complexed ligand is found.

[0113] Next, the three dimensional geometric structure of the active site is determined. This can be done by known methods, including X-ray crystallography, which can determine a complete molecular structure. On the other hand, solid or liquid phase NMR can be used to determine certain intramolecular distances. Any other experimental method of structure determination can be used to obtain partial or complete geometric structures. The geometric structures may be measured with a complexed ligand, natural or artificial, which may increase the accuracy of the active site structure determined.

[0114] If an incomplete or insufficiently accurate structure is determined, the methods of computer based numerical modeling can be used to complete the structure or improve its accuracy. Any recognized modeling method may be used, including parameterized models specific to particular biopolymers such as proteins or nucleic acids, molecular dynamics models based on computing molecular motions, statistical mechanics models based on thermal ensembles, or combined models. For most types of models, standard molecular force fields, representing the forces between constituent atoms and groups, are necessary, and can be selected from force fields known in physical chemistry. The incomplete or less accurate experimental structures can serve as constraints on the complete and more accurate structures computed by these modeling methods.

[0115] Finally, having determined the structure of the active site, either experimentally, by modeling, or by a combination, candidate modulating compounds can be identified by searching databases containing compounds along with information on their molecular structure. Such a search seeks compounds having structures that match the determined active site structure and that interact with the groups defining the active site. Such a search can be manual, but is preferably computer assisted. These compounds found from this search are potential NGPCR modulating compounds.

[0116] Alternatively, these methods can be used to identify improved modulating compounds from an already known modulating compound or ligand. The composition of the known compound can be modified and the structural effects of modification can be determined using the experimental

and computer modeling methods described above applied to the new composition. The altered structure is then compared to the active site structure of the compound to determine if an improved fit or interaction results. In this manner systematic variations in composition, such as by varying side groups, can be quickly evaluated to obtain modified modulating compounds or ligands of improved specificity or activity.

[0117] Further experimental and computer modeling methods useful to identify modulating compounds based upon identification of the active sites of a NGPCR, and related transduction and transcription factors will be apparent to those of skill in the art.

[0118] Examples of molecular modeling systems are the CHARMM and QUANTA programs (Polygen Corporation, Waltham, MA). CHARMM performs the energy minimization and molecular dynamics functions. QUANTA performs the construction, graphic modeling and analysis of molecular structure. QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

[0119] A number of articles review computer modeling of drugs interactive with specific proteins, such as Rotivinen, *et al.*, 1988, *Acta Pharmaceutical Fennica* 97:159-166; Ripka, *New Scientist* 54-57 (June 16, 1988); McKinaly and Rossmann, 1989, *Annu. Rev. Pharmacol. Toxicol.* 29:111-122; Perry and Davies, *OSAR: Quantitative Structure-Activity Relationships in Drug Design* pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis and Dean, 1989 *Proc. R. Soc. Lond.* 236:125-140 and 141-162; and, with respect to a model receptor for nucleic acid components, Askew, *et al.*, 1989,

J. Am. Chem. Soc. 111:1082-1090. Other computer programs that screen and graphically depict chemicals are available from companies such as BioDesign, Inc. (Pasadena, CA.), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario). Although these are primarily designed for application to drugs specific to particular proteins, they can be adapted to design of drugs specific to regions of DNA or RNA, once that region is identified.

[0120] Although described above with reference to design and generation of compounds which could alter binding, one could also screen libraries of known compounds, including natural products or synthetic chemicals, and biologically active materials, including proteins, for compounds which are inhibitors or activators.

[0121] Cell-based systems can also be used to identify compounds that bind NGPCRs as well as assess the altered activity associated with such binding in living cells. Assays for agonists and antagonists of NGPCRS that can be used in cell-based systems according to certain embodiments include, but are not limited to, those de-scribed in U.S. Patent Serial No. 6,004,808, and PCT Application Number US99/17425, which are herein incorporated by reference in their entirety for any purpose.

[0122] One tool of particular interest for such assays is green fluorescent protein which is described, *inter alia*, in U.S. Patent No. 5,625,048, --herein incorporated by reference. Cells that may be used in such cellular assays include, but are not limited to, leukocytes, or cell lines derived from leukocytes, lymphocytes, stem cells, including embryonic stem cells, and the

like. In addition, expression host cells (e.g., B95 cells, COS cells, CHO cells, OMK cells, fibroblasts, Sf9 cells) genetically engineered to express a functional NGPCR of interest and to respond to activation by the test, or natural, ligand, as measured by a chemical or phenotypic change, or induction of another host cell gene, can be used as an end point in the assay.

[0123] Compounds identified via assays such as those described herein may be useful, for example, in elaborating certain biological functions of a NGPCR gene product. Such compounds can be administered to a patient at therapeutically effective doses to treat any of a variety of physiological or mental disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in any amelioration, impediment, prevention, or alteration of any biological or overt symptom.

[0124] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub>/ED<sub>50</sub>.

Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, typically care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.



[0125] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0126] Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, intracranial, intrathecal, or rectal administration.

[0127] For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl

methycellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

[0128] Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

[0129] For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

[0130] For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other

suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0131] The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[0132] The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

[0133] In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion

exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

[0134] The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

IN VITRO SCREENING ASSAYS FOR COMPOUNDS THAT BIND TO  
NGPCRs

[0135] In vitro systems may be designed to identify compounds capable of interacting with (e.g., binding to) NGPCR (including, but not limited to, a ECD or CD of NGPCR). Compounds identified may be useful, for example, in modulating the activity of wild type and/or mutant NGPCR gene products; may be useful in elaborating certain biological functions of the NGPCR; may be utilized in screens for identifying compounds that disrupt normal NGPCR interactions; or may in themselves disrupt such interactions.

[0136] The principle of the assays used to identify compounds that bind to the NGPCR involves preparing a reaction mixture of the NGPCR and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex which can be removed and/or detected in the reaction mixture. The NGPCR species used can vary depending upon the goal of the screening assay. For example, in certain embodiments, where agonists of the natural ligand are sought, the full length NGPCR, or a soluble truncated NGPCR, e.g., in which the TM and/or CD is deleted from the molecule, a peptide corresponding to a ECD or a

fusion protein containing one or more NGPCR ECD fused to a protein or polypeptide that affords advantages in the assay system (e.g., labeling, isolation of the resulting complex, etc.) can be utilized. Where compounds that interact with the cytoplasmic domain are sought to be identified, peptides corresponding to the NGPCR CD and fusion proteins containing the NGPCR CD can be used.

[0137] The screening assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the NGPCR protein, polypeptide, peptide or fusion protein or the test substance onto a solid phase and detecting NGPCR/test compound complexes anchored on the solid phase at the end of the reaction. In certain embodiments of such a method, the NGPCR reactant may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

[0138] In practice, in certain embodiments, microtiter plates may conveniently be utilized as the solid phase. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished by simply coating the solid surface with a solution of the protein and drying. In certain embodiments, an immobilized antibody, e.g., a monoclonal antibody, specific for the protein to be immobilized may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

[0139] In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. In certain

embodiments, after the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. In various embodiments, the detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, in certain embodiments, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

[0140] In certain embodiments, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for a NGPCR protein, polypeptide, peptide or fusion protein or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

[0141] In certain embodiments, cell-based assays can be used to identify compounds that interact with NGPCR. To this end, cell lines that express NGPCR, or cell lines (e.g., COS cells, CHO cells, fibroblasts, etc.) that have been genetically engineered to express a NGPCR (e.g., by transfection or transduction of NGPCR DNA) can be used. Interaction of the

test compound with, for example, a ECD of a NGPCR expressed by the host cell can be determined by comparison or competition with native ligand.

#### 5.5.2. ASSAYS FOR INTRACELLULAR PROTEINS THAT INTERACT WITH NGPCRs

[0142] Any method suitable for detecting protein-protein interactions may be employed for identifying transmembrane proteins or intracellular proteins that interact with a NGPCR. Among the methods which may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns of cell lysates or proteins obtained from cell lysates and a NGPCR to identify proteins in the lysate that interact with the NGPCR. For these assays, the NGPCR component used can be a full length NGPCR, a soluble derivative lacking the membrane-anchoring region (e.g., a truncated NGPCR in which a TM is deleted resulting in a truncated molecule containing a ECD fused to a CD), a peptide corresponding to a CD or a fusion protein containing a CD of a NGPCR. Once isolated, such an intracellular protein can be identified and can, in turn, be used, in conjunction with standard techniques, to identify proteins with which it interacts. For example, at least a portion of the amino acid sequence of an intracellular protein which interacts with a NGPCR can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique. (See, e.g., Creighton, 1983, "Proteins: Structures and Molecular Principles", W.H. Freeman & Co., N.Y., pp.34-49). The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences

encoding such intracellular proteins. Screening can be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (See, e.g., Ausubel, *supra*, and PCR Protocols: A Guide to Methods and Applications, 1990, Innis, M. et al., eds. Academic Press, Inc., New York).

[0143] Additionally, methods may be employed which result in the simultaneous identification of genes which encode the transmembrane or intracellular proteins interacting with NGPCR. These methods include, for example, probing expression, libraries, in a manner similar to the well known technique of antibody probing of  $\lambda$ gt11 libraries, using labeled NGPCR protein, or an NGPCR polypeptide, peptide or fusion protein, e.g., an NGPCR polypeptide or NGPCR domain fused to a marker (e.g., an enzyme, fluor, luminescent protein, or dye), or an Ig-Fc domain.

[0144] One method that detects protein interactions in vivo, the two-hybrid system, is described in detail for illustration only and not by way of limitation. One version of this system has been described (Chien et al., 1991, Proc. Natl. Acad. Sci. USA, 88:9578-9582) and is commercially available from Clontech (Palo Alto, CA).

[0145] Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one plasmid has nucleotides encoding the DNA-binding domain of a transcription activator protein fused to a NGPCR nucleotide sequence encoding NGPCR, an NGPCR polypeptide, peptide or fusion protein, and the other plasmid has nucleotides encoding the transcription activator protein's activation domain fused to a cDNA encoding



an unknown protein which has been recombined into this plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., HBS or lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene: the DNA-binding domain hybrid cannot because it does not provide activation function and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

[0146] The two-hybrid system or related methodology may be used to screen activation domain libraries for proteins that interact with the "bait" gene product. By way of example, and not by way of limitation, a NGPCR may be used as the bait gene product. Total genomic or cDNA sequences are fused to the DNA encoding an activation domain. This library and a plasmid encoding a hybrid of a bait NGPCR gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, a bait NGPCR gene sequence, such as the open reading frame of a NGPCR (or a domain of a NGPCR) can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated.

DNA sequencing is then used to identify the proteins encoded by the library plasmids.

[0147] A cDNA library of the cell line from which proteins that interact with bait NGPCR gene product are to be detected can be made using methods routinely practiced in the art. According to certain embodiments of the system described herein, for example, the cDNA fragments can be inserted into a vector such that they are translationally fused to the transcriptional activation domain of GAL4. This library can be co-transformed along with the bait NGPCR gene-GAL4 fusion plasmid into a yeast strain which contains a lacZ gene driven by a promoter which contains GAL4 activation sequence. A cDNA encoded protein, fused to GAL4 transcriptional activation domain, that interacts with bait NGPCR gene product will reconstitute an active GAL4 protein and thereby drive expression of the HIS3 gene. Colonies which express HIS3 can be detected by their growth on petri dishes containing semi-solid agar based media lacking histidine. The cDNA can then be purified from these strains, and used to produce and isolate the bait NGPCR gene-interacting protein using techniques routinely practiced in the art.

ASSAYS FOR COMPOUNDS THAT INTERFERE WITH  
NGPCR/INTRACELLULAR OR NGPCR/TRANSMEMBRANE  
MACROMOLECULE INTERACTION

[0148] The macromolecules that interact with the NGPCR are referred to, for purposes of this discussion, as "binding partners." These binding partners are likely to be involved in the NGPCR signal transduction pathway. Therefore, it is desirable to identify compounds that interfere with or disrupt

the interaction of such binding partners which may be useful in regulating the activity of a NGPCR and controlling disorders associated with NGPCR activity. For example, given their expression pattern, the described NGPCRs are contemplated to be particularly useful in methods for identifying compounds useful in the therapeutic treatment of obesity, inflammation, immune disorders, diabetes, heart and coronary disease, metabolic disorders, and cancer.

[0149] In certain embodiments, assay systems used to identify compounds that interfere with the interaction between a NGPCR and its binding partner or partners involve preparing a reaction mixture containing NGPCR protein, polypeptide, peptide or fusion protein as described herein, and the binding partner under conditions and for a time sufficient to allow the two to interact and bind, thus forming a complex. In order to test a compound for inhibitory activity, the reaction mixture is prepared in the presence and absence of the test compound. The test compound may be initially included in the reaction mixture, or may be added at a time subsequent to the addition of the NGPCR moiety and its binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the NGPCR moiety and the binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the NGPCR and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal NGPCR protein may also be compared to

complex formation within reaction mixtures containing the test compound and a mutant NGPCR. This comparison may be important in those cases wherein it is desirable to identify compounds that specifically disrupt interactions of mutant, or mutated, NGPCRs but not normal NGPCRs.

[0150] According to certain embodiments, the assay for compounds that interfere with the interaction of a NGPCR and its binding partners can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the NGPCR moiety product or the binding partner onto a solid phase and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction by competition can be identified by conducting the reaction in the presence of the test substance; i.e., by adding the test substance to the reaction mixture prior to, or simultaneously with, a NGPCR moiety and interactive binding partner. Alternatively, test compounds that disrupt preformed complexes, e.g. compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. Various formats, according to certain embodiments, are described briefly below.

[0151] In a heterogeneous assay system, either a NGPCR moiety or an interactive binding partner, is anchored onto a solid surface, while the non-anchored species is labeled, either directly or indirectly. In practice, microtiter

plates are conveniently utilized. The anchored species may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the NGPCR gene product or binding partner and drying. Alternatively, an immobilized antibody specific for the species to be anchored may be used to anchor the species to the solid surface. The surfaces may be prepared in advance and stored.

[0152] To conduct the assay according to certain embodiments, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the non-immobilized species is pre-labeled, in certain embodiments, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, in certain embodiments, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds which inhibit complex formation or which disrupt preformed complexes can be detected.

[0153] In certain embodiments, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction

products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds which inhibit complex or which disrupt preformed complexes can be identified.

[0154] In certain embodiments of the invention, a homogeneous assay can be used in which a preformed complex of a NGPCR moiety and an interactive binding partner is prepared in which either the NGPCR or its binding partner is labeled, but the signal generated by the label is quenched due to formation of the complex (see, e.g., U.S. Patent No. 4,109,496 by Rubenstein which utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances which disrupt NGPCR/intracellular binding partner interaction can be identified.

[0155] In certain embodiments, a NGPCR fusion can be prepared for immobilization. For example, a NGPCR or a peptide fragment, e.g., corresponding to a CD, can be fused to a glutathione-S-transferase (GST) gene using a fusion vector, such as pGEX-5X-1, in such a manner that its binding activity is maintained in the resulting fusion protein. The interactive binding partner can be purified and used to raise a monoclonal antibody, using methods routinely practiced in the art and described herein. This

antibody can be labeled with the radioactive isotope  $^{125}\text{I}$ , for example, by methods routinely practiced in the art. In a heterogeneous assay, e.g., the GST-NGPCR fusion protein can be anchored to glutathione-agarose beads. The interactive binding partner can then be added in the presence or absence of the test compound in a manner that allows interaction and binding to occur. At the end of the reaction period, unbound material can be washed away, and the labeled monoclonal antibody can be added to the system and allowed to bind to the complexed components. The interaction between a NGPCR gene product and the interactive binding partner can be detected by measuring the amount of radioactivity that remains associated with the glutathione-agarose beads. A successful inhibition of the interaction by the test compound will result in a decrease in measured radioactivity.

[0156] In certain embodiments, the GST-NGPCR fusion protein and the interactive binding partner can be mixed together in liquid in the absence of the solid glutathione-agarose beads. The test compound can be added either during or after the species are allowed to interact. This mixture can then be added to the glutathione-agarose beads and unbound material is washed away. Again the extent of inhibition of the NGPCR/binding partner interaction can be detected by adding the labeled antibody and measuring the radioactivity associated with the beads.

[0157] In certain embodiments of the invention, these same techniques can be employed using peptide fragments that correspond to the binding domains of a NGPCR and/or the interactive or binding partner (in cases where the binding partner is a protein), in place of one or both of the full

length proteins. Any number of methods routinely practiced in the art can be used to identify and isolate the binding sites. These methods include, but are not limited to, mutagenesis of the gene encoding one of the proteins and screening for disruption of binding in a co-immunoprecipitation assay.

Compensatory mutations in the gene encoding the second species in the complex can then be selected. Sequence analysis of the genes encoding the respective proteins will reveal the mutations that correspond to the region of the protein involved in interactive binding. Alternatively, one protein can be anchored to a solid surface using methods described above, and allowed to interact with and bind to its labeled binding partner, which has been treated with a proteolytic enzyme, such as trypsin. After washing, a relatively short, labeled peptide comprising the binding domain may remain associated with the solid material, which can be isolated and identified by amino acid sequencing. Also, once the gene coding for the intracellular binding partner is obtained, short gene segments can be engineered to express peptide fragments of the protein, which can then be tested for binding activity and purified or synthesized.

[0158] For example, and not by way of limitation, a NGPCR gene product can be anchored to a solid material as described, above, by making a GST-NGPCR fusion protein and allowing it to bind to glutathione agarose beads. The interactive binding partner can be labeled with a radioactive isotope, such as  $^{35}\text{S}$ , and cleaved with a proteolytic enzyme such as trypsin. Cleavage products can then be added to the anchored GST-NGPCR fusion protein and allowed to bind. After washing away unbound peptides, labeled



bound material, representing the intracellular binding partner binding domain, can be eluted, purified, and analyzed for amino acid sequence by well-known methods. Peptides so identified can be produced synthetically or fused to appropriate facilitative proteins using recombinant DNA technology.

[0159] The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as illustrations of individual aspects of certain embodiments of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims. All referenced documents, including publications, patents, and patent applications, are incorporated by reference herein for any purpose.

## WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence selected from:
  - a) the nucleotide sequence as set forth in SEQ ID NO: 1;
  - b) the nucleotide sequence as set forth in SEQ ID NO: 3;
  - c) the nucleotide sequence as set forth in SEQ ID NO: 5; and
  - d) the nucleotide sequence as set forth in SEQ ID NO: 6.
2. The isolated nucleic acid molecule of claim 1, comprising a nucleotide sequence as set forth in SEQ ID NO: 1.
3. An isolated nucleic acid molecule comprising at least 22 contiguous bases of a nucleotide sequence selected from:
  - a) the nucleotide sequence as set forth in SEQ ID NO: 1;
  - b) the nucleotide sequence as set forth in SEQ ID NO: 3;
  - c) the nucleotide sequence as set forth in SEQ ID NO: 5; and
  - d) the nucleotide sequence as set forth in SEQ ID NO: 6.
4. An isolated nucleic acid molecule comprising a nucleotide sequence that is at least 95% identical to a nucleotide sequence selected from:
  - a) the nucleotide sequence as set forth in SEQ ID NO: 1;
  - b) the nucleotide sequence as set forth in SEQ ID NO: 3;
  - c) the nucleotide sequence as set forth in SEQ ID NO: 5; and
  - d) the nucleotide sequence as set forth in SEQ ID NO: 6.

5. An isolated nucleic acid molecule comprising a nucleotide sequence encoding the amino acid sequence as set forth in SEQ ID NO: 2.
6. An isolated nucleic acid molecule comprising a nucleotide sequence encoding the amino acid sequence as set forth in SEQ ID NO: 4.
7. An isolated nucleic acid molecule comprising a nucleotide sequence encoding the amino acid sequence as set forth in SEQ ID NO: 7.
8. An isolated polypeptide comprising an amino acid sequence selected from:
  - a) the amino acid sequence as set forth in SEQ ID NO: 2;
  - b) the amino acid sequence as set forth in SEQ ID NO: 4;and
  - c) the amino acid sequence as set forth in SEQ ID NO: 7.
9. The isolated polypeptide of claim 8, comprising an amino acid sequence as set forth in SEQ ID NO: 2.
10. An isolated polypeptide comprising an amino acid sequence that is at least 95% identical to an amino acid sequence selected from:
  - a) the amino acid sequence as set forth in SEQ ID NO: 2;
  - b) the amino acid sequence as set forth in SEQ ID NO: 4;and
  - c) the amino acid sequence as set forth in SEQ ID NO: 7.
11. An isolated polypeptide comprising an amino acid sequence selected from:

- a) the amino acid sequence as set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution;
  - b) the amino acid sequence as set forth in SEQ ID NO: 4 with at least one conservative amino acid substitution; and
  - c) the amino acid sequence as set forth in SEQ ID NO: 7 with at least one conservative amino acid substitution.
12. An isolated polypeptide of claim 11, comprising an amino acid sequence as set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution.
13. A recombinant host cell containing an isolated nucleic acid molecule of any of claims 1, 2, 3, 4, 5, or 6.
14. A recombinant host cell of claim 13, wherein the host cell is eukaryotic.
15. A recombinant host cell of claim 14, wherein the host cell is selected from CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, and WI38 cell lines.
16. A method for identifying a molecule that binds to the polypeptide of one of claims 8, 10, or 11, comprising:
- (a) contacting the polypeptide with the molecule in vitro; and
  - (b) detecting the binding of said polypeptide to said molecule.

## SEQUENCE LISTING

<110> Lexicon Genetics Incorporated

<120> Novel Seven Transmembrane Proteins and Polynucleotides Encoding the Same

<130> 7705.15-304

<140> Not Yet Assigned

<141> 2001-05-11

<150> 60/203,875

<151> 2000-05-12

<150> 60/207,932

<151> 2000-05-30

<160> 7

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2985

<212> DNA

<213> homo sapiens

<400> 1

atggtctgtt	cggtgcccc	actgctgctc	ctggccacaa	ctcttccct	gctgggggtca	60
ccagttgccc	aagcatccca	acctggacag	agtcaggctg	gaggggaatc	tggatctggg	120
cagctcctgg	accaagagaa	tggagcaggg	gaatcagcgc	tggctccgt	ctatgtacat	180
ctggactttc	cagataagac	ctggccccct	gaactctcca	ggacactgac	tctccctgct	240
gcctcagctt	cctcttcccc	aaggcctctt	ctcaactggc	tcagactcac	aacagagtgt	300
aatgtcaacc	acaaggggaa	tttctattgt	gcttgccctc	ctggctacca	gtggaacacc	360
agcatctgoc	tccattaccc	tccttgtcaa	agcctccaca	accaccagcc	ttgtggctgc	420
cttgtcttca	gccatcccga	acccgggtac	tgccagttgc	tgccacctgt	ccccgggatc	480
ctcaacctga	actcccagct	gcagatgcct	ggtgacacgc	tgagcctgac	tctccatctg	540
agccaggagg	ccaccaacct	gagctggttc	ctgaggcacc	cagggagccc	cagtcccatc	600
ctcctgcagc	cagggacaca	ggtgtctgtg	acttccagcc	acggccaggc	tgccctcagc	660
gtctccaaca	tgtcccatca	ctgggcaggt	gagtacatga	gctgcttoga	ggcccagggc	720
ttcaagtgga	acctgtatga	ggtggtgagg	gtgcccttga	aggcgacaga	tgtggctoga	780
cttccatacc	agctgtccat	ctcctgtgcc	acctcccttg	gcttccagct	gagctgctgc	840
atccccagca	caaacctggc	ctacaccgcg	gcctggagcc	ctggagaggg	cagcaaagct	900
tcctccttca	acgagtcagg	ctctcagtgc	tttgtgctgg	ctgttcagcg	ctgcccgatg	960
gctgacacca	cgtacacttg	tgacctgcag	agcctgggccc	tggtccact	caggggtccc	1020
atctccatca	ccatcatcca	ggatggagac	atcacctgcc	ctgaggacgc	ctcgggtgctc	1080
acctggaatg	tcaccaaggc	tggccacgtg	gcacaggccc	catgtcctga	gagcaagagg	1140
ggcatagtga	ggaggctctg	tggggctgac	ggagtctggg	ggccgggtcca	cagcagctgc	1200
acagatgcga	ggctcctggc	cttgttcact	agaaccaagc	tgctgcaggc	aggccagggc	1260
agtctgctg	aggaggtgcc	acagatcctg	gcacagctgc	cagggcaggc	ggcagaggca	1320
agttcaccc	ccgacttact	gaccctgctg	agcaccatga	aatacgtggc	caaggtggtg	1380
gcagaggcca	gaatacagct	tgaccgcaga	gccctgaaga	atctcctgat	tgccacagac	1440
aaggtcctag	atatggacac	caggtctctg	tggaccttgg	ccaagccc	gaagccctgg	1500
gcaggctcga	ctctcctgct	ggctgtggag	accctggcat	gcagcctgtg	cccacaggac	1560

```

tacccttctg ccttcagctt acccaatgtg ctgctgcaga gccagctgtt tggacccacg 1620
tttcctgctg actacagcat ctccttccct actcgcccc cactgcaggc tcagattccc 1680
aggcactcac tggccccatt ggtccgtaat ggaactgaaa taagtattac tagcctggtg 1740
ctgcgaaaac tggaccacct tctgccccca aactatggac aagggtgtgg ggattccctc 1800
tatgccactc ctggcctggt ccttgtcatt tccatcatgg caggtgaccg ggccttcagc 1860
caggagaggg tcatcatgga ctttggaac acagatggtt cccctcactg tgccttctgg 1920
gatcacagtc tcttcaggg cagggggggg tgggtccaaag aagggtgcca ggcacaggtg 1980
gccagtgcca gcccactgc tcagtgcctc tgccagcacc tcaactgcctt ctccgtcctc 2040
atgtccccc acactgttcc ggaagaaccc gctctggcgc tgctgactca rgtgggcttg 2100
ggagcttcca tactggcgct gcttgtgtgc ctgggtgtgt actggctggt gtggagagtc 2160
gtggtgcgga acaagatctc ctatttccgc cagcccgccc tgctcaacat ggtgttctgc 2220
ttgctggccg cagacacttg cttcctgggc gcccatttcc tctctccagg gccccgaagc 2280
ccgctctgcc ttgctgcgc cttcctctgt catttctct acctggccac ctttttctgg 2340
atgctggcg aggccctggt gttggccac cagctgctct ttgtctttca ccagctggca 2400
aagcaccgag ttctccccct catggtgctc ctgggctacc tgtgccact ggggttgga 2460
ggtgtcacc tggggtctc cctacctca gggcaatacc tgaggaggag ggaatgctg 2520
ttggatgga agggaggggc gttatacacc ttctggtggc cagtgtgtgc catcataggc 2580
gtgaatggc tggtagtag catggccatg ctgaagtgtc tgagacctc gctgtcagag 2640
ggacccccag cagagaagcg ccaagctctg ctgggggtga tcaaagccct gctcattctt 2700
acacccatct ttggcctcac ctgggggtg ggcctggcca ctctgttaga ggaagtctcc 2760
acggtccctc attacatctt caccattctc aacacccctc agggcgtctt catcctattg 2820
tttggttgcc tcatggacag gaagatacaa gaagctttgc gcaaacgctt ctgccgcgcc 2880
caagcccca gctccaccat ctccctggcc acaaataga gctgcatctt ggaacacagc 2940
aaaggaggaa gcgacactgc caggaagaca gatgcttcag agtga 2985

```

<210> 2  
 <211> 994  
 <212> PRT  
 <213> homo sapiens

<400> 2

Met	Val	Cys	Ser	Ala	Ala	Pro	Leu	Leu	Leu	Leu	Ala	Thr	Thr	Leu	Pro
1				5				10						15	
Leu	Leu	Gly	Ser	Pro	Val	Ala	Gln	Ala	Ser	Gln	Pro	Gly	Gln	Ser	Gln
			20				25						30		
Ala	Gly	Gly	Glu	Ser	Gly	Ser	Gly	Gln	Leu	Leu	Asp	Gln	Glu	Asn	Gly
		35					40					45			
Ala	Gly	Glu	Ser	Ala	Leu	Val	Ser	Val	Tyr	Val	His	Leu	Asp	Phe	Pro
	50				55						60				
Asp	Lys	Thr	Trp	Pro	Pro	Glu	Leu	Ser	Arg	Thr	Leu	Thr	Leu	Pro	Ala
65				70					75					80	
Ala	Ser	Ala	Ser	Ser	Ser	Pro	Arg	Pro	Leu	Leu	Thr	Gly	Leu	Arg	Leu
			85					90					95		
Thr	Thr	Glu	Cys	Asn	Val	Asn	His	Lys	Gly	Asn	Phe	Tyr	Cys	Ala	Cys
			100				105						110		
Leu	Ser	Gly	Tyr	Gln	Trp	Asn	Thr	Ser	Ile	Cys	Leu	His	Tyr	Pro	Pro
		115				120						125			
Cys	Gln	Ser	Leu	His	Asn	His	Gln	Pro	Cys	Gly	Cys	Leu	Val	Phe	Ser
	130				135						140				
His	Pro	Glu	Pro	Gly	Tyr	Cys	Gln	Leu	Leu	Pro	Pro	Val	Pro	Gly	Ile
145				150					155					160	
Leu	Asn	Leu	Asn	Ser	Gln	Leu	Gln	Met	Pro	Gly	Asp	Thr	Leu	Ser	Leu
			165				170						175		
Thr	Leu	His	Leu	Ser	Gln	Glu	Ala	Thr	Asn	Leu	Ser	Trp	Phe	Leu	Arg
		180					185						190		
His	Pro	Gly	Ser	Pro	Ser	Pro	Ile	Leu	Leu	Gln	Pro	Gly	Thr	Gln	Val
	195					200						205			
Ser	Val	Thr	Ser	Ser	His	Gly	Gln	Ala	Ala	Leu	Ser	Val	Ser	Asn	Met

210		215		220
Ser His His Trp Ala Gly	Glu Tyr Met Ser Cys Phe Glu Ala Gln Gly			
225	230	235	240	
Phe Lys Trp Asn Leu Tyr Glu Val Val Arg Val Pro Leu Lys Ala Thr				
	245	250	255	
Asp Val Ala Arg Leu Pro Tyr Gln Leu Ser Ile Ser Cys Ala Thr Ser				
	260	265	270	
Pro Gly Phe Gln Leu Ser Cys Cys Ile Pro Ser Thr Asn Leu Ala Tyr				
	275	280	285	
Thr Ala Ala Trp Ser Pro Gly Glu Gly Ser Lys Ala Ser Ser Phe Asn				
	290	295	300	
Glu Ser Gly Ser Gln Cys Phe Val Leu Ala Val Gln Arg Cys Pro Met				
305	310	315	320	
Ala Asp Thr Thr Tyr Thr Cys Asp Leu Gln Ser Leu Gly Leu Ala Pro				
	325	330	335	
Leu Arg Val Pro Ile Ser Ile Thr Ile Ile Gln Asp Gly Asp Ile Thr				
	340	345	350	
Cys Pro Glu Asp Ala Ser Val Leu Thr Trp Asn Val Thr Lys Ala Gly				
	355	360	365	
His Val Ala Gln Ala Pro Cys Pro Glu Ser Lys Arg Gly Ile Val Arg				
	370	375	380	
Arg Leu Cys Gly Ala Asp Gly Val Trp Gly Pro Val His Ser Ser Cys				
385	390	395	400	
Thr Asp Ala Arg Leu Leu Ala Leu Phe Thr Arg Thr Lys Leu Leu Gln				
	405	410	415	
Ala Gly Gln Gly Ser Pro Ala Glu Glu Val Pro Gln Ile Leu Ala Gln				
	420	425	430	
Leu Pro Gly Gln Ala Ala Glu Ala Ser Ser Pro Ser Asp Leu Leu Thr				
	435	440	445	
Leu Leu Ser Thr Met Lys Tyr Val Ala Lys Val Val Ala Glu Ala Arg				
	450	455	460	
Ile Gln Leu Asp Arg Arg Ala Leu Lys Asn Leu Leu Ile Ala Thr Asp				
465	470	475	480	
Lys Val Leu Asp Met Asp Thr Arg Ser Leu Trp Thr Leu Ala Gln Ala				
	485	490	495	
Arg Lys Pro Trp Ala Gly Ser Thr Leu Leu Leu Ala Val Glu Thr Leu				
	500	505	510	
Ala Cys Ser Leu Cys Pro Gln Asp Tyr Pro Phe Ala Phe Ser Leu Pro				
	515	520	525	
Asn Val Leu Leu Gln Ser Gln Leu Phe Gly Pro Thr Phe Pro Ala Asp				
	530	535	540	
Tyr Ser Ile Ser Phe Pro Thr Arg Pro Pro Leu Gln Ala Gln Ile Pro				
545	550	555	560	
Arg His Ser Leu Ala Pro Leu Val Arg Asn Gly Thr Glu Ile Ser Ile				
	565	570	575	
Thr Ser Leu Val Leu Arg Lys Leu Asp His Leu Leu Pro Ser Asn Tyr				
	580	585	590	
Gly Gln Gly Leu Gly Asp Ser Leu Tyr Ala Thr Pro Gly Leu Val Leu				
	595	600	605	
Val Ile Ser Ile Met Ala Gly Asp Arg Ala Phe Ser Gln Gly Glu Val				
	610	615	620	
Ile Met Asp Phe Gly Asn Thr Asp Gly Ser Pro His Cys Val Phe Trp				
625	630	635	640	
Asp His Ser Leu Phe Gln Gly Arg Gly Gly Trp Ser Lys Glu Gly Cys				
	645	650	655	
Gln Ala Gln Val Ala Ser Ala Ser Pro Thr Ala Gln Cys Leu Cys Gln				
	660	665	670	
His Leu Thr Ala Phe Ser Val Leu Met Ser Pro His Thr Val Pro Glu				

	675		680		685
Glu Pro	Ala Leu Ala Leu	Leu Thr Gln Val Gly	Leu Gly Ala Ser Ile		
690		695	700		
Leu Ala	Leu Leu Val Cys	Leu Gly Val Tyr Trp	Leu Val Trp Arg Val		
705		710	715		720
Val Val	Arg Asn Lys Ile	Ser Tyr Phe Arg His	Ala Ala Leu Leu Asn		
	725		730		735
Met Val	Phe Cys Leu Leu	Ala Ala Asp Thr Cys	Phe Leu Gly Ala Pro		
	740		745		750
Phe Leu	Ser Pro Gly Pro	Arg Ser Pro Leu	Cys Leu Ala Ala Phe		
	755		760		765
Leu Cys	His Phe Leu Tyr	Leu Ala Thr Phe Phe	Trp Met Leu Ala Gln		
	770		775		780
Ala Leu	Val Leu Ala His	Gln Leu Leu Phe Val	Phe His Gln Leu Ala		
785		790	795		800
Lys His	Arg Val Leu Pro	Leu Met Val Leu	Leu Gly Tyr Leu Cys	Pro	
	805		810		815
Leu Gly	Leu Ala Gly Val	Thr Leu Gly Leu Tyr	Leu Pro Gln Gly Gln		
	820		825		830
Tyr Leu	Arg Glu Gly Glu	Cys Trp Leu Asp Gly	Lys Gly Gly Ala Leu		
	835		840		845
Tyr Thr	Phe Val Gly Pro	Val Leu Ala Ile Ile	Gly Val Asn Gly Leu		
	850		855		860
Val Leu	Ala Met Ala Met	Leu Lys Leu Leu Arg	Pro Ser Leu Ser Glu		
865		870	875		880
Gly Pro	Pro Ala Glu Lys	Arg Gln Ala Leu	Leu Gly Val Ile Lys	Ala	
	885		890		895
Leu Leu	Ile Leu Thr Pro	Ile Phe Gly Leu Thr	Trp Gly Leu Gly Leu		
	900		905		910
Ala Thr	Leu Leu Glu Glu	Val Ser Thr Val Pro	His Tyr Ile Phe Thr		
	915		920		925
Ile Leu	Asn Thr Leu Gln	Gly Val Phe Ile Leu	Leu Phe Gly Cys Leu		
	930		935		940
Met Asp	Arg Lys Ile Gln	Glu Ala Leu Arg	Lys Arg Phe Cys Arg	Ala	
945		950	955		960
Gln Ala	Pro Ser Ser Thr	Ile Ser Leu Ala	Thr Asn Glu Gly Cys	Ile	
	965		970		975
Leu Glu	His Ser Lys Gly	Gly Gly Ser Asp	Thr Ala Arg Lys Thr	Asp Ala	
	980		985		990
Ser Glu					

&lt;210&gt; 3

&lt;211&gt; 2481

&lt;212&gt; DNA

&lt;213&gt; homo sapiens

&lt;400&gt; 3

atgcctgggtg	acacgctgag	cctgactctc	catctgagcc	aggaggccac	caacctgagc	60
tggttcctga	ggcaccagg	gagccccagt	cccatcctcc	tgcagccagg	gacacaggtg	120
tctgtgactt	ccagccacgg	ccaggctgcc	ctcagcgtct	ccaacatgtc	ccatcactgg	180
gcaggtgagt	acatgagctg	cttcgaggcc	cagggcttca	agtggaacct	gtatgaggtg	240
gtgagggtgc	ccttgaaggc	gacagatgtg	gctcgacttc	cataccagct	gtccatctcc	300
tgtgccacct	cccctggctt	ccagctgagc	tgctgcatcc	ccagcacaaa	cctggcctac	360
accgcggcct	ggagccctgg	agagggcagc	aaagcttcct	ccttcaacga	gtcaggctct	420
cagtgccttg	tgctggctgt	tcagcgctgc	ccgatggctg	acaccacgta	cacttgtgac	480
ctgcagagcc	tgggcctggc	tccactcagg	gtccccatct	ccatcaccat	catccaggat	540
ggagacatca	cctgccctga	ggacgcctcg	gtgctcacct	ggaatgtcac	caaggctggc	600
cacgtggcac	aggccccatg	tcctgagagc	aagaggggca	tagtgaggag	gctctgtggg	660



```

gctgacggag tctggggggc ggtccacagc agctgcacag atgcgaggct cctggccttg 720
ttcactagaa ccaagctgct gcaggcaggc cagggcagtc ctgctgagga ggtgccacag 780
atcctggcac agctgccagg gcaggcggca gaggcaagtt caccctccga cttactgacc 840
ctgctgagca ccatgaaata cgtggccaag gtggtggcag aggccagaat acagcttgac 900
cgcagagccc tgaagaatct cctgattgcc acagacaagg tcctagatat ggacaccagg 960
tctctgtgga ccctggccca agcccggaa ccctgggcag gctcgactct cctgctggct 1020
gtggagaccc tggcatgcag cctgtgcccc caggactacc ccttcgcctt cagcttacct 1080
aatgtgctgc tgcagagcca gctgttttga cccacgtttc ctgctgacta cagcatctcc 1140
ttccctactc ggccccact gcaggctcag attcccaggc actcactggc cccattggctc 1200
cgtaatggaa ctgaaataag tattactagc ctggtgctgc gaaaactgga ccaccttctg 1260
ccctcaaaact atggacaagg gctgggggat tccctctatg ccactcctgg cctggctcctt 1320
gtcatttcca tcatggcagg tgaccggggc ttccagccagg gagaggctcat catggacttt 1380
gggaacacag atggttcccc tcaactgtgc ttctgggata acagtctctt ccagggcagg 1440
gggggttggg ccaaagaagg gtgccaggca caggtggcca gtgccagccc cactgctcag 1500
tgctctgcc agcacctcac tgccttctcc gtctcatgt cccacacac tgttccggaa 1560
gaacccgctc tggcgctgct gactcargtg ggcttgggag cttccatact ggcgctgctt 1620
gtgtgcctgg gtgtgtactg gctggtgtgg agagtctgtg tgcggaacaa gatctcctat 1680
ttccgccacg ccgccctgct caacatggtg ttctgcttgc tggccgcaga cacttgcttc 1740
ctgggcgccc cattcctctc tccaggcccc cgaagccgcg tctgccttgc tgccgccttc 1800
ctctgtcatt tcctctacct ggccaccttt ttctggatgc tggcgcaggc cctgggtgtt 1860
gcccaccagc tgctctttgt ctttcaccag ctggcacaagc accagattct cccctcatg 1920
gtgctcctgg gctacctgtg cccactgggg ttggcaggtg tcaccctggg gctctaccta 1980
cctcaagggg aatacctgag ggagggggaa tgctggttgg atgggaaggg aggggcgtta 2040
tacaccttcg tggggccagt gctggccatc ataggcgtga atgggctggg actagccatg 2100
gccatgctga agttgctgag accttcgctg tcagagggac cccagcaga gaagcgcaa 2160
gctctgctgg gggatgatcaa agccctgctc attcttacac ccattcttgg cctcacctgg 2220
gggctggggc tggccactct gttagaggaa gtctccacgg tccctcatta catcttcacc 2280
attctcaaca ccctccaggg cgtcttcac cttattgtttg gttgcctcat ggacaggaag 2340
atacaagaag ctttgcgcaa acgcttctgc cgcgcccaag ccccagctc caccatctcc 2400
ctggccacaa atgaaggctg catcttgaa cacagcaaag gaggaagcga cactgccagg 2460
aagacagatg cttcagagtg a 2481

```

```

<210> 4
<211> 826
<212> PRT
<213> homo sapiens

```

```

<400> 4
Met Pro Gly Asp Thr Leu Ser Leu Thr Leu His Leu Ser Gln Glu Ala
1 5 10 15
Thr Asn Leu Ser Trp Phe Leu Arg His Pro Gly Ser Pro Ser Pro Ile
20 25 30
Leu Leu Gln Pro Gly Thr Gln Val Ser Val Thr Ser Ser His Gly Gln
35 40 45
Ala Ala Leu Ser Val Ser Asn Met Ser His His Trp Ala Gly Glu Tyr
50 55 60
Met Ser Cys Phe Glu Ala Gln Gly Phe Lys Trp Asn Leu Tyr Glu Val
65 70 75 80
Val Arg Val Pro Leu Lys Ala Thr Asp Val Ala Arg Leu Pro Tyr Gln
85 90 95
Leu Ser Ile Ser Cys Ala Thr Ser Pro Gly Phe Gln Leu Ser Cys Cys
100 105 110
Ile Pro Ser Thr Asn Leu Ala Tyr Thr Ala Ala Trp Ser Pro Gly Glu
115 120 125
Gly Ser Lys Ala Ser Ser Phe Asn Glu Ser Gly Ser Gln Cys Phe Val
130 135 140
Leu Ala Val Gln Arg Cys Pro Met Ala Asp Thr Thr Tyr Thr Cys Asp
145 150 155 160

```

```

Leu Gln Ser Leu Gly Leu Ala Pro Leu Arg Val Pro Ile Ser Ile Thr
      165      170      175
Ile Ile Gln Asp Gly Asp Ile Thr Cys Pro Glu Asp Ala Ser Val Leu
      180      185      190
Thr Trp Asn Val Thr Lys Ala Gly His Val Ala Gln Ala Pro Cys Pro
      195      200      205
Glu Ser Lys Arg Gly Ile Val Arg Arg Leu Cys Gly Ala Asp Gly Val
      210      215      220
Trp Gly Pro Val His Ser Ser Cys Thr Asp Ala Arg Leu Leu Ala Leu
      225      230      235      240
Phe Thr Arg Thr Lys Leu Leu Gln Ala Gly Gln Gly Ser Pro Ala Glu
      245      250      255
Glu Val Pro Gln Ile Leu Ala Gln Leu Pro Gly Gln Ala Ala Glu Ala
      260      265      270
Ser Ser Pro Ser Asp Leu Leu Thr Leu Leu Ser Thr Met Lys Tyr Val
      275      280      285
Ala Lys Val Val Ala Glu Ala Arg Ile Gln Leu Asp Arg Arg Ala Leu
      290      295      300
Lys Asn Leu Leu Ile Ala Thr Asp Lys Val Leu Asp Met Asp Thr Arg
      305      310      315      320
Ser Leu Trp Thr Leu Ala Gln Ala Arg Lys Pro Trp Ala Gly Ser Thr
      325      330      335
Leu Leu Leu Ala Val Glu Thr Leu Ala Cys Ser Leu Cys Pro Gln Asp
      340      345      350
Tyr Pro Phe Ala Phe Ser Leu Pro Asn Val Leu Leu Gln Ser Gln Leu
      355      360      365
Phe Gly Pro Thr Phe Pro Ala Asp Tyr Ser Ile Ser Phe Pro Thr Arg
      370      375      380
Pro Pro Leu Gln Ala Gln Ile Pro Arg His Ser Leu Ala Pro Leu Val
      385      390      395      400
Arg Asn Gly Thr Glu Ile Ser Ile Thr Ser Leu Val Leu Arg Lys Leu
      405      410      415
Asp His Leu Leu Pro Ser Asn Tyr Gly Gln Gly Leu Gly Asp Ser Leu
      420      425      430
Tyr Ala Thr Pro Gly Leu Val Leu Val Ile Ser Ile Met Ala Gly Asp
      435      440      445
Arg Ala Phe Ser Gln Gly Glu Val Ile Met Asp Phe Gly Asn Thr Asp
      450      455      460
Gly Ser Pro His Cys Val Phe Trp Asp His Ser Leu Phe Gln Gly Arg
      465      470      475      480
Gly Gly Trp Ser Lys Glu Gly Cys Gln Ala Gln Val Ala Ser Ala Ser
      485      490      495
Pro Thr Ala Gln Cys Leu Cys Gln His Leu Thr Ala Phe Ser Val Leu
      500      505      510
Met Ser Pro His Thr Val Pro Glu Pro Ala Leu Ala Leu Leu Thr
      515      520      525
Gln Val Gly Leu Gly Ala Ser Ile Leu Ala Leu Leu Val Cys Leu Gly
      530      535      540
Val Tyr Trp Leu Val Trp Arg Val Val Val Arg Asn Lys Ile Ser Tyr
      545      550      555      560
Phe Arg His Ala Ala Leu Leu Asn Met Val Phe Cys Leu Leu Ala Ala
      565      570      575
Asp Thr Cys Phe Leu Gly Ala Pro Phe Leu Ser Pro Gly Pro Arg Ser
      580      585      590
Pro Leu Cys Leu Ala Ala Ala Phe Leu Cys His Phe Leu Tyr Leu Ala
      595      600      605
Thr Phe Phe Trp Met Leu Ala Gln Ala Leu Val Leu Ala His Gln Leu
      610      615      620

```

Leu Phe Val Phe His Gln Leu Ala Lys His Arg Val Leu Pro Leu Met  
 625 630 635 640  
 Val Leu Leu Gly Tyr Leu Cys Pro Leu Gly Leu Ala Gly Val Thr Leu  
 645 650 655  
 Gly Leu Tyr Leu Pro Gln Gly Gln Tyr Leu Arg Glu Gly Glu Cys Trp  
 660 665 670  
 Leu Asp Gly Lys Gly Gly Ala Leu Tyr Thr Phe Val Gly Pro Val Leu  
 675 680 685  
 Ala Ile Ile Gly Val Asn Gly Leu Val Leu Ala Met Ala Met Leu Lys  
 690 695 700  
 Leu Leu Arg Pro Ser Leu Ser Glu Gly Pro Pro Ala Glu Lys Arg Gln  
 705 710 715 720  
 Ala Leu Leu Gly Val Ile Lys Ala Leu Leu Ile Leu Thr Pro Ile Phe  
 725 730 735  
 Gly Leu Thr Trp Gly Leu Gly Leu Ala Thr Leu Leu Glu Glu Val Ser  
 740 745 750  
 Thr Val Pro His Tyr Ile Phe Thr Ile Leu Asn Thr Leu Gln Gly Val  
 755 760 765  
 Phe Ile Leu Leu Phe Gly Cys Leu Met Asp Arg Lys Ile Gln Glu Ala  
 770 775 780  
 Leu Arg Lys Arg Phe Cys Arg Ala Gln Ala Pro Ser Ser Thr Ile Ser  
 785 790 795 800  
 Leu Ala Thr Asn Glu Gly Cys Ile Leu Glu His Ser Lys Gly Gly Ser  
 805 810 815  
 Asp Thr Ala Arg Lys Thr Asp Ala Ser Glu  
 820 825

&lt;210&gt; 5

&lt;211&gt; 3207

&lt;212&gt; DNA

&lt;213&gt; homo sapiens

&lt;400&gt; 5

tgcagttctt	ccttgcggaa	aatttgcac	gtgggcttgt	agggactgtt	tcattgagcc	60
atggtctgtt	cggctgcccc	actgctgctc	ctggccacaa	ctcttcccct	gctgggggtca	120
ccagttgccc	aagcatccca	acctggacag	agtcaggctg	gaggggaatc	tggatctggg	180
cagctcctgg	accaagagaa	tggagcaggg	gaatcagcgc	tggctccgt	ctatgtacat	240
ctggactttc	cagataagac	ctggccccct	gaactctcca	ggacactgac	tctcccctgct	300
gcctcagctt	cctcttcccc	aaggcctctt	ctcactggcc	tcagactcac	aacagagtgt	360
aatgtcaacc	acaaggggaa	tttctattgt	gcttgccctc	ctggctacca	gtggaacacc	420
agcatctgcc	tccattaccc	tccttgtcaa	agcctccaca	accaccagcc	ttgtggctgc	480
cttgtcttca	gccatcccga	acccgggtac	tgccagttgc	tgccacctgt	ccccgggatc	540
ctcaacctga	actcccagct	gcagatgcct	ggtgacacgc	tgagcctgac	tctccatctg	600
agccaggagg	ccaccaacct	gagctggttc	ctgaggcacc	cagggagccc	cagtcccatc	660
ctcctgcagc	cagggacaca	ggtgtctgtg	acttccagcc	acggccaggc	tgccctcagc	720
gtctccaaca	tgtcccatca	ctgggcaggt	gagtacatga	gctgcttcga	ggcccagggc	780
ttcaagtgga	acctgtatga	ggtggtgagg	gtgcccttga	aggcgacaga	tgtggctcga	840
cttccatacc	agctgtccat	ctcctgtgcc	acctcccctg	gcttccagct	gagctgctgc	900
atccccagca	caaacctggc	ctacaccgcg	gcctggagcc	ctggagaggg	cagcaaagct	960
tctctcttca	acgagtcagg	ctctcagtgc	tttgtgctgg	ctgttcagcg	ctgcccgatg	1020
gctgacacca	cgtacacttg	tgacctgcag	agcctggggc	tggctccact	caggggtccc	1080
atctccatca	ccatcatcca	ggatggagac	atcacctgcc	ctgaggacgc	ctcgggtgctc	1140
acctggaatg	tcaccaaggc	tggccacgtg	gcacaggccc	catgtcctga	gagcaagagg	1200
ggcatagtga	ggaggctctg	tggggctgac	ggagtctggg	ggccgggtcca	cagcagctgc	1260
acagatgcga	ggctcctggc	cttgttcact	agaaccaagc	tgctgcaggc	aggccagggc	1320
agtctgctg	aggaggtgcc	acagatcctg	gcacagctgc	cagggcaggc	ggcagaggca	1380
agttcacctt	ccgacttact	gacctgtctg	agcaccatga	aatacgtggc	caaggtgggtg	1440
gcagaggcca	gaatacagct	tgaccgcaga	gccctgaaga	atctcctgat	tgccacagac	1500

aaggtcctag	atatggacac	caggtctctg	tggaccctgg	cccaagcccg	gaagccctgg	1560
gcaggctcga	ctctcctgct	ggctgtggag	accctggcat	gcagcctgtg	cccacaggac	1620
taccccttcg	ccttcagctt	acccaatgtg	ctgctgcaga	gccagctgtt	tggacccacg	1680
tttcctgctg	actacagcat	ctccttcctt	actcggcccc	cactgcaggc	tcagattccc	1740
aggcactcac	tggccccatt	ggtccgtaat	ggaactgaaa	taagtattac	tagcctgggtg	1800
ctgcgaaaac	tggaccacct	tctgccctca	aactatggac	aagggtctgg	ggattccctc	1860
tatgccactc	ctggcctggg	ccttgtcatt	tccatcatgg	caggtgaccg	ggccttcagc	1920
cagggagagg	tcacatcatgga	ctttgggaac	acagatgggt	cccctcactg	tgtcttctgg	1980
gatcacagtc	tcttcagggt	cagggggggg	tgggtccaaag	aagggtgcca	ggcacagggtg	2040
gccagtgcc	gccccactgc	tcagtgcctc	tgccagcacc	tcactgcctt	ctccgtcctc	2100
atgtccccac	acactgttcc	ggaagaaccc	gctctggcgc	tgtctgactca	rgtgggcttg	2160
ggagcttcca	tactggcgct	gcttgtgtgc	ctgggtgtgt	actggctggg	gtggagagtc	2220
gtggtgcgga	acaagatctc	ctatttccgc	cacgccgccc	tgtctaacat	ggtgttctgc	2280
ttgctggccg	cagacacttg	cttcctgggc	gccccattcc	tctctccagg	gccccgaagc	2340
ccgctctgcc	ttgctgcgcg	cttcctctgt	catttccctc	acctggccac	cttttctctg	2400
atgctggcgc	aggccctggg	gttggccccc	cagctgctct	ttgtctttca	ccagctggca	2460
aagcaccgag	ttctccccct	catggtgtct	ctgggtacc	tgtgcccact	ggggttggca	2520
ggtgtcacc	tggggctcta	cctacctcaa	gggcaatacc	tgaggaggag	ggaatgctgg	2580
ttggatggga	agggaggggc	gttatacacc	ttcgtggggc	cagtgtctgg	catcataggc	2640
gtgaatgggc	tgggtactagc	catggccatg	ctgaagttgc	tgagaccttc	gctgtcagag	2700
ggacccccag	cagagaagcg	ccaagctctg	ctgggggtga	tcaaagccct	gctcattctt	2760
acacccatct	ttggcctcac	ctgggggctg	ggcctggcca	ctctgttaga	ggaagtctcc	2820
acggtccctc	attacattct	caccattctc	aacacctcc	agggcgtctt	catcctattg	2880
tttggttgcc	tcattggacag	gaagatacaa	gaagctttgc	gcaaacgctt	ctgccgcgcc	2940
caagccccca	gctccaccat	ctccctggcc	acaaatgaag	gctgcatctt	ggaacacagc	3000
aaaggaggaa	gcgacactgc	caggaagaca	gatgcttcag	agtgaaccac	acacggaccc	3060
atgttcctgc	aagggagttg	aggctgtgtg	cttgaacca	ccagatgagc	cctggcccaa	3120
tgctctgaac	tcttcccgcc	tcccggagct	cagcccttga	gaaaggcagg	cttatatttc	3180
ccttagtgac	actcatttat	cttacag				3207

&lt;210&gt; 6

&lt;211&gt; 1008

&lt;212&gt; DNA

&lt;213&gt; homo sapiens

&lt;400&gt; 6

atgcagccgt	ccccgcgcgc	caccgagctg	gtgccgtcgg	agcgcgccgt	ggtgctgctg	60
tcgtgcgcac	tctccgcgct	cggctcgggc	ctgctgggtg	ccacgcacgc	cctgtggccc	120
gacctgcgca	gccgggcacg	gcgcctgctg	ctcttcctgt	cgttgccga	cctgctctcg	180
gccgcctcct	acttctacgg	agtgtgcag	aacttcgcg	gccgctcgtg	ggaactgcgtg	240
ctgcaggcgg	cgtgtgccac	cttcgccaac	accagctcct	tcttctggac	cgtggccatt	300
gcgctctact	tgtacctcag	catcgtccgc	gccgcgcgcg	ggcctcgcac	agatcgccctg	360
ctttgggcct	tccatgtcgt	cagctggggg	gtcccgttgg	tcactcactgt	ggcagccgtc	420
gccctgaaga	agattggcta	tgacgcctcg	gacgtgtctg	tgggctgggtg	ctggatcgac	480
ctggaggcca	aggaccatgt	cctgtggatg	ctgctgacgg	ggaagctgtg	ggagatgctg	540
gcataatgtgc	tgtctgcctct	gctgtacctc	ctgggtccgga	agcacatcaa	cagagcgcac	600
acggcactct	ctgagtaccg	gcccatacctc	tcccaggagc	accgcctgct	gcgccactcc	660
tccatggcgg	acaagaagct	ggtgctcatc	ccgctcatct	tcactcggcct	cagggtctgg	720
agcaccgtgc	ggttcgtgct	gacctctgt	ggctccccgg	ccgtgcagac	gccggtgctg	780
gtggttctgc	atggtatcgg	gaacacgttt	caggggagggtg	ccaactgcat	catgttctgc	840
ctctgcaccc	gcgcgcgtccg	aactcggctc	ttctctctct	gttgtgctg	ctgctcttct	900
cagcctccca	ccaagagccc	ggctggcact	cccaaggctc	ccgcgccttc	caagccagga	960
gaatgtcagg	aatcccaagg	gacccagggg	gaacttccaa	gcacctga		1008

&lt;210&gt; 7

&lt;211&gt; 335

&lt;212&gt; PRT

&lt;213&gt; homo sapiens

&lt;400&gt; 7

```

Met Gln Pro Ser Pro Pro Pro Thr Glu Leu Val Pro Ser Glu Arg Ala
 1          5          10          15
Val Val Leu Leu Ser Cys Ala Leu Ser Ala Leu Gly Ser Gly Leu Leu
          20          25          30
Val Ala Thr His Ala Leu Trp Pro Asp Leu Arg Ser Arg Ala Arg Arg
          35          40          45
Leu Leu Leu Phe Leu Ser Leu Ala Asp Leu Leu Ser Ala Ala Ser Tyr
 50          55          60
Phe Tyr Gly Val Leu Gln Asn Phe Ala Gly Pro Ser Trp Asp Cys Val
65          70          75          80
Leu Gln Gly Ala Leu Ser Thr Phe Ala Asn Thr Ser Ser Phe Phe Trp
          85          90          95
Thr Val Ala Ile Ala Leu Tyr Leu Tyr Leu Ser Ile Val Arg Ala Ala
          100          105          110
Arg Gly Pro Arg Thr Asp Arg Leu Leu Trp Ala Phe His Val Val Ser
          115          120          125
Trp Gly Val Pro Leu Val Ile Thr Val Ala Ala Val Ala Leu Lys Lys
130          135          140
Ile Gly Tyr Asp Ala Ser Asp Val Ser Val Gly Trp Cys Trp Ile Asp
145          150          155          160
Leu Glu Ala Lys Asp His Val Leu Trp Met Leu Leu Thr Gly Lys Leu
          165          170          175
Trp Glu Met Leu Ala Tyr Val Leu Leu Pro Leu Leu Tyr Leu Leu Val
          180          185          190
Arg Lys His Ile Asn Arg Ala His Thr Ala Leu Ser Glu Tyr Arg Pro
          195          200          205
Ile Leu Ser Gln Glu His Arg Leu Leu Arg His Ser Ser Met Ala Asp
210          215          220
Lys Lys Leu Val Leu Ile Pro Leu Ile Phe Ile Gly Leu Arg Val Trp
225          230          235          240
Ser Thr Val Arg Phe Val Leu Thr Leu Cys Gly Ser Pro Ala Val Gln
          245          250          255
Thr Pro Val Leu Val Val Leu His Gly Ile Gly Asn Thr Phe Gln Gly
          260          265          270
Gly Ala Asn Cys Ile Met Phe Val Leu Cys Thr Arg Ala Val Arg Thr
          275          280          285
Arg Leu Phe Ser Leu Cys Cys Cys Cys Cys Ser Ser Gln Pro Pro Thr
290          295          300
Lys Ser Pro Ala Gly Thr Pro Lys Ala Pro Ala Pro Ser Lys Pro Gly
305          310          315          320
Glu Ser Gln Glu Ser Gln Gly Thr Pro Gly Glu Leu Pro Ser Thr
          325          330          335

```